OM protein - protein search, using sw model

Run on: January 25, 2000, 03:59:51 ; Search time 29.27 Seconds

GenCore version 4.5
Copyright (c) 1993 - 1996 Compugen Ltd.

Title: US-09-332-063-2
Perfect score: 3347
Sequence: 1 MPRAQPSSASYQVPADPP\..........KTPIQILGQEPDAEMVEYLI 675

Scoring table: BLOSUM62

Searched: 188963 seqs, 23686106 residues
Da ase :   A_Geneseq_35:*

Word size : 0

Number of hits that pass the threshold :   188963

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	229.5	6.9	783	1	W37151	Mouse neural Mena+
2	229	6.9	802	1	W37153	Mouse neural Mena+
3	228.5	6.8	787	1	W37152	Mouse neural Mena+
4	226.5	6.8	787	1	R72826	Mouse neural Mena+
5	226.5	6.8	2482	1	W29996	Human mitosin. Pur
6	225	6.7	3248	1	R99795	Human mitosin amin
7	216	6.5	442	1	W26540	Kinetochore protei
8	213	6.4	1972	1	W00024	Trypanosoma cruzi
9	213	6.4	262	1	W06913	Smooth muscle myos
10	212.5	6.3	1274	1	W89253	T. cruzi L19E homo
11	211	6.3	1411	1	W02258	Human ALP. New nuc
12	209.5	6.3	1315	1	W76734	Nucleolar/endosoma
-	207.5	6.2	561	1	W63043	Human mDia Rho tar
14	205.5	6.2	1885	1	W54241	Streptococcus uber
15	204	6.1	2272	1	W21731	Rattus norvegicus
16	204	6.1	2192	1	W21732	GAL4/HA/NuMA fusio
17	203	6.1	180	1	R22392	LexA/NuMA fusion p
18	203	6.1	576	1	R66929	Antigen tc-7a. Vac
19	203	6.1	885	1	R66930	AML chromosome in
20	203	6.1	816	1	R66931	AML chromosome in
21	201.5	6.0	2101	1	R47173	Sequence of the in
22	199.5	6.0	1388	1	R56475	Protein with Rho p
23	199	5.9	1255	1	W52249	Mouse Rho target p
24	199	5.9	1255	1	W76733	Mouse mDia Rho tar
25	198	5.9	1201	1	W90345	Drosophila sp. Cos
26	197.5	5.9	1372	1	W56473	Protein with Rho p
27	197	5.9	3119	1	W72204	HSV-2 strain SB5 C
28	196.5	5.8	1392	1	Y05999	Restin protein seq
29	194	5.8	1404	1	R26049	MSF precursor. New
30	194	5.8	206	1	W14574	Streptococcus pneu
31	193.5	5.8	955	1	R57305	K39 polypeptide of
32	193.5	5.8	1427	1	W03651	Leishmania chagasi
33	193.5	5.7	1093	1	R10534	Human 160kD mediat
34	191	5.7	1093	1	R42818	TMF. New protein c
35	190.5	5.7	407	1	W94500	M. tuberculosis RP
36	190	5.7	765	1	W31852	Mycobacterium tube
37	190	5.7	572	1	W31855	Mycobacterium tube
38	188.5	5.6	700	1	W52622	Human NonO/TFE3 fu
39	188	5.6	929	1	W14593	Streptococcus pneu

ALIGNMENTS

RESULT 1

ID   W37151 standard; Protein; 783 AA.
AC   W37151;
DT   06-JUL-1998   (first entry)
DE   Mouse neural Mena+ protein.
KW   Neural Mena+ protein; mammalian Ena; Enabled protein; Evl protein;
KW   cytoskeleton; cell, morphology; cell adhesion; cell differentiation;
KW   cell growth; cell motility; mouse.
OS   Mus musculus.
OS   Mus musculus.
FH   Key   Location/Qualifiers
FH
FT   Misc.difference 378   /note= "encoded by GG?"
PN   WO9601/5-A1.
PD   16-JAN-1998.
PF   ?9-JUL-1997; U11669.
PR   05-JUL-1996; US-675815.
PA   (UBF3) GES BIOTECHNOLOGISCHE FORSCHUNG MBH.
PA   (HICC-) HUTCHINSON CANCER RES CENT FRED.
PI   Gertler FB, Niebuhr K, Soriano P, Wehland J;
DR   N-PSDB; V02998
WPI; 98-101197/09.

Detection of modulators of Mena and Ena-VASP-like genes and proteins
- used in control of cytoskeletal dynamic events in normal and
abnormal cell morphology; cell morphology, adhesion, motility, growth and
differentiation
Example 4; Page 58-60; 77pp; English.
This protein comprises novel murine neural Mena+.   Its amino acid
sequence was deduced from a cDNA clone (see V02998) obtained from
a mouse brain cDNA library. Neural Mena+ contains an exon that
introduces 244 amino acids between amino acids 238 and 239 of
mammalian Ena (Mena, see W37148). Two other isoforms, neural
Mena++ (see W37152) and neural Mena+++ (see W37153), are also
disclosed. Unlike Mena, neural Mena isoforms exhibit neural
tissue-specific distribution. Based on the disclosed Mena and Evl
genes (see also V02996-97) and proteins (see also W37148-49), a
variety of methods and compositions are provided for screening,
isolating and characterising endogenous and exogenous factors,
drugs and therapeutic agents useful to evaluate and/or control
cytoskeletal dynamic events involved in normal and abnormal cell
morphology, adhesion, motility, growth and/or differentiation. A
method of detecting a modulator of Mena activity/expression is
claimed.

Query Match   6.9%;   Score 229.5;   DB 1;   Length 783;
Best Local Similarity   20.7%;   Pred. No. 1.5e-06;
Matches   155;   Conservative   65;   Mismatches   205;   Indels   221;   Gaps   27;

Qy   98 FNFDLPRLETANKQLAEKEYEGSEJTRKITSQLFA--KNKESQREKEKLEAELATARST 155
              | | | : ||| ||| | || || | || | :||
Db   109 FASVMTIALEVLNSQ------EAGPTLPRONSQLPAQVQVQNGPSOEELEIQRRQL---- 148

Qy   156 NEDLXRHIEIRDQALSNAQAKVVGLEELKKKKQVVDKVEKMQQALVQLQAACEKREQLE 215
              | || :| | ||| | | : |||| ||| | :| :|| |||
Db   148 -QECOROKELEREIMERERLEREJPLEPEPLER----ERLEQE--QLERQOREHVE 197

Qy   216 SRLRVLERELESLRIQQRQGNCPQTNVSFYNAAALMELLREKEKERILALEADMTKWEQK 275
              || || | || | | | || ||| || | :| |||: | ||
Db   198 RLPPLRER-LERER-QERE-- -------------------RERLEQLEREOVEWERE 232

RESULT 2

W37153

ID   W37153 standard; Protein; 802 AA.
AC   W37153;
DT   06-JUL-1998 (first entry)
DE   Mouse neural Mena+++ protein.
KW   Neural Mena+++ protein; mammalian Ena; Enabled protein; Evl protein;
KW   cytoskeleton; cell morphology; cell adhesion; cell differentiation;
KW   cell growth; cell motility; mouse.
OS   Mus musculus.
PN   WO9801755-A1.
PD   15-JAN-1998.
PF   03-JUL-1997; U11669.
PR   05-JUL-1996; US-675815.
PA   (GBFB) GES BIOTECHNOLOGISCHE FORSCHUNG MBH.
PA   (HUTC-) HUTCHINSON CANCER RES CENT FRED.
PI   ertler FB, Niebuhr K, Soriano P, Wehland J;
DR   WPI; 98-101197/09.
PT   - used in control of cytoskeletal dynamic events in normal and
PT   abnormal cell morphology, adhesion, motility, growth and
PT   differentiation
PS   Example 4; Page 63-65; 77pp; English.
PT   This protein comprises novel murine neural Mena+++. Its amino acid
PT   sequence was deduced from a cDNA clone obtained from a mouse brain
PT   cDNA library. Two other isoforms, neural Mena+ (see W37151) and
PT   neural Mena++ (see W37152), are also disclosed. Unlike mammalian
CC   Ena (Mena, see W37148), neural Mena isoforms exhibit neural
CC   tissue-specific distribution. Based on the disclosed Mena and Evl
CC   genes (see also V02996-98) and proteins (see also W37148-49), a
CC   variety of methods and compositions are provided for screening,
CC   isolating and characterising endogenous and exogenous factors,
CC   drugs and therapeutic agents useful to evaluate and/or control
CC   cytoskeletal dynamic events involved in normal and abnormal cell
CC   morphology, adhesion, motility, growth and/or differentiation. A
CC   method of detecting a modulator of Mena activity/expression is
CC   claimed.
SQ   Sequence 802 AA;

```
Qy    276 YLEENVMRHFALDAAATVAAQRDTTVISHSPNTSYDTALEARIQKEEEILMANKRCLDM 335
              |:||:|||||||| ||   :||  |     ||||   |||  |  :|||
Db    233 RRMSNAAPSSDSSLSSAPLPEYSSCCOPPSAPPSYAKVISAPVSDAIPDY-------- 283

Qy    336 EGRIKTLHAQIIEKDAMIKVLQQRSREPSKTEQLSCMRPA--KSLMSISNAGSGLLSHS 393
                     :::           |  |  |    :| ||  |:  |
Db    283 ------AVVTAL-------PPTSTPTPPLRHAATRPATSLGSSAFHPVLPHY 321

Qy    394 STLTGSPIMEEKRDDKSWKGSLGILLGGDYRAEYVPSTP------------------ 433
              |:|
Db    322 ATVP-RPLNKNSR------------PSSPVNTPSSQPPRAAKSCAWPTSNF 358

Qy    433 SPVPPSTPLLSAH--SKTGSRDCSTQTERGTESNKTAAVAPISVPAPVAAAATAAAITA 489
              |:||||||   :||  ||  |  :| ||       |  ||
Db    359 SPLPPSPPIMISSPPGKATGPR------------------PVLPVCVSSPVPQMPPSP--TA 400

Qy    490 TAAITITTMVAAAPAVAVAAAAP--------------AAAAAPSP 520
                :||  |||  |     :|
Db    401 PNGSLDSVTYPVSPPTTSGPAAPPPPPPPPPPPPPLPPPPLPPLASLSHCGSQOASPPP 460

Qy    521 ATAAATAAAVSPAAAGQIPAAASVASAAAAVAP------------SAAAAAVQVAPAAPAPVPAP 573
              :| |      |     ::|         |    |:|  |
Db    461 GTPLASTPSSKPSV---LPSPSAGAPASAETPLNPELGDSSASEPGLQAA-SQPAESPPP 516

Qy    574 -ALV---PVPAPAAAQASAPAQOTQAPTSAPAVAPTPAPTPTPAVAQAEVPASPATGPGPH 629
              |:|                  ||  ||  |  |:|      |
Db    517 QGLVLGPPAPPPPPLPSGPAY------ASALPPPGPPPPPPL------PSTGPPPP 562

Qy    630 RLSIPSLTCNPDKTDGP------------VFHSNTLERKTPI 659
                     |    :||  |:  ||   :|:
Db    563 PPPPPPL--PNQAPPPPPPPAPPLPASGIFSGSTSEDNRPL 602
```

Query Match   6.8%; Score 229; DB 1; Length 802;
Best Local Similarity 20.4%;
Matches 134; Conservative 89; Mismatches 205; Indels 228; Gaps 28;
                                Pred. No. 1.7e-05;

```
Qy     98 FNRDLRERLETANKQ-LAEKEYEGSEDTR-----------------KTISQLFA--KNKESQREK 142
             |:                              :|  |    |  |:
Db    100 FASAMMHALEVLNSQCEAAQSKVTAFIQDSTNLRCIFCGPTLPRONSQLPAQVQNGPSQEEL 159

Qy    143 EKLEAELATARSTNEDQCRRHIEIRDQALSNAQAKVVKLEEELKKKQVYVDKVERMQQALV 202
              |:  |            :||: :|  | | |
Db    160 EIQRRQL-------QEOQQROKELERERMERERLERERLERER------ERLEQE-- 204

Qy    203 QLQAACEKREQLEHRLRTRLERELESLRIQOROGNCQOPTNVSEYNAAALMELLREKEERI 262
              || |  ||||  |||:  ||||  ||
Db    204 QLERQRQERSHVERLERERLER-LERER-QERE------------------RERL 238

Qy    263 LALEADMTKWEQKYLEENVMRHFALDAAATVAAQRDTTVISHSPNTSYDTALEARIQKEE 322
              :|  |  |:|::||||||||||||||||||||||||||||||||||||||||||||||||
Db    239 EQLEREQVEWERERRMSNAAPSSDSSLSSAPLPEYSSCQPPSAPPSYAKVISAPVSDAT 298

Qy    323 EEILMANKRCLDMEGRIKTLHAQIIEKDAMIKVLQQRSREPSKTEQLSCMRPA--KSLM 380
                                 :::             |  |    :| ||  |:
Db    299 PDY----------------AVVTAL------PPTSTPTPPLRHAATRFAT 327

Qy    381 SISNAGSGLLSHSSTLTGSPIMEEKRDDKSWKGSLGILLGGDYRAEYVPSTP------- 433
                              |:|
Db    328 SIGSAFHPVLPHYATVP-RPLNKNSR---------------PSSPVNTPSSQP 364

Qy    433 ------SPVPPSTPLLSAH---SKTGSRDCSTQTERGTESNKTAAVAPISVPA 476
                     |:|||||   |||   ||  |  :|
Db    365 PAAKSCAWPTSNFSPLPPSPPIMISSPPGKATGPR--------------PVLPVCVSS 408

Qy    477 PVAAAATAAAITATAAITITTTMVAAAPAVAVAAAAP--------------- 513
                  :||  |||  |
Db    409 PVPQMPPSP--TAPNGSLDSVTPVSPPTTSGPAAPPPPPPPPPPPPLPPPPLPPLA 466

Qy    513 -----AAAAAPSPATAAATAAAVSPAAAGQIPAAASVASAAAAV------SAAAAAV 560
                          :| |      |     ::|         |
Db    467 SLSHCGSQOASPPGTPLASTPSSKPSV---LPSPSAGAPASAETPLNPELGDSSASEPGL 523

Qy    561 QVAPAAPAPVPAP-ALV---PVPAPAAAQASAPAQOTQAPTSAPAVAPTPAPTPTPAVAQA 616
                    |    |:|                  ||  ||  |
Db    524 OAA-SQPAESPPTPQGLVLGPPAPPPPPLPSGPAY-----ASALPPPGPPPPPPL--- 574

Qy    617 EVPASPATGPGPHRLSIPSLTCNPDKTDGP------------VFHSNTLERKTPI 659
                     |    |:  ||   :|:
Db    574 -----PSTGPPPPPPPPPPL---PNQAPPPPPPPAPPLPASGIFSGSTSEDNRPL 621
```

RESULT 3

W37152

ID   W37152 standard; Protein; 787 AA.
AC   W37152;
DT   06-JUL-1998 (first entry)
DE   Mouse neural Mena++ protein.
KW   Neural Mena++ protein; mammalian Ena; Enabled protein; Evl protein;
KW   cytoskeleton; cell morphology; cell adhesion; cell differentiation;
KW   cell growth; cell motility; mouse.
OS   Mus musculus.
PN   WO9801755-A1.
PD   15-JAN-1998.
PF   03-JUL-1997; U11669.
PR   05-JUL-1996; US-675815.
PA   (GBFB) GES BIOTECHNOLOGISCHE FORSCHUNG MBH.
PA   (HUTC-) HUTCHINSON CANCER RES CENT FRED.
PI   Gertler FB, Niebuhr K, Soriano P, Wehland J;
DR   WPI; 98-101197/09.
PT   Detection of modulators of Mena and Ena-VASP-like genes and proteins
PT   - used in control of cytoskeletal dynamic events in normal and
PT   abnormal cell morphology, adhesion, motility, growth and
PT   differentiation
PS   Example 4; Page 60-63; 77pp; English.
CC   This protein comprises novel murine neural Mena++. Its amino acid

```
CC   sequence was deduced from a cDNA clone obtained from a mouse brain
CC   cDNA library. Two other isoforms, neural Mena+ (see W37151) and
CC   neural Mena+++ (see W37153), are also disclosed. Unlike mammalian
CC   Ena (Mena, see W37148), neural Mena isoforms exhibit neural
CC   tissue-specific distribution. Based on the disclosed Mena and Evl
CC   genes (see also V02996-98) and proteins (see also W37148-49), a
CC   variety of methods and compositions are provided for screening,
CC   isolating and characterising endogenous and exogenous factors,
CC   drugs and therapeutic agents useful to evaluate and/or control
CC   cytoskeletal dynamic events involved in normal and abnormal cell
CC   morphology, adhesion, motility, growth and/or differentiation. A
CC   method of detecting a modulator of Mena activity/expression is
CC   claimed.
SQ   Sequence   787 AA;
```

```
Query Match             6.8%;   Score 228.5;   DB 1;   Length 787;
P-t Local Similarity     21.2%;  Pred. No. 1.8e-06;
hes 136; Conservative   84;   Mismatches 206;   Indels 217;   Gaps   28;

Qy    98 FNRDLRERLETANKQLAEKEYEGSEDTRKTISQLFA--KNKESQREKEKLEAELATARST 155
         ::|:|| : ||: :| : ||| |: |||  :  || |
Db   100 FASAMMHALEVLNSQEA-VFYLGPTLPRQN-SQLPAQVQNGPSPQELEIQRRQL------ 152

Qy   156 NEDQRRHIEIRDQALSNAQAKVVKLEEELIKKKQVVDKYEKMQQALVQLQAACEKREQLE 215
         :||| :   |: |||:| : | ::||||||||  | | |||||||||||||||||||||
Db   152 -QEQORQKELEREMRERLERERERLER------ ERLEQE--QLERQREHVE 201

Qy   216 HRLRTRLERELESLRIQORGNCQPTNVSEYNAAALMELLREKEERILALEADMTKWEQK 275
         || | |||||| |||||||||||||||||||||||||||||||||||||||||||||||
Db   202 RLERERLER-LERER-QERE------                    -RERLEQLEREQVEWERE 236

Qy   276 YLEENVMRHFALDAAATVAAQRDTTVISHSPNTSYDTALEARIQKEEEEILMANKRCLDM 335
         :||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   237 RRMSNAAPSDSSLLSSAPLPEYSSCQPSAPPPSYAKVISAPVSDATPDY------ 287

Qy   336 EGRIKTLHAQIIEKDAMIKVLQQRSRKEPSKTEQLSCMRPA--KSLMSISNAGSGLLSHS 393
         |::                       ::|                :||::::| |:| |:|
Db   287 --------------AVVTAL---PFTSTPPTPLRHAATRFATSLGSAFHPVLPHY 325

Qy   394 STLTGSPIMEEKRDDKSWKGSLGILLLGGDYRAEYVPSTP----- 433
         |: :|                        ||:| |||| |
Db   326 ATVP-RPLNKNSR-----                     PSSPVNTPSSQPPAAKSCAWPTSNF 362

Qy   433 SPVPPSTPFLLSAH--SKTGSRDCSTQTERGTESNKTAAVAPISVPAPVAAAATAAATTA 489
         |:|||: ||||||   |: |: | | : | : ||  |:| |:  ||  |||||| |:
Db   363 SPLPPSPPIMISSPPGKATGPR-----                     -PVLPVCVSSPVPQMPFSP--TA 404

Qy   490 TAATITTTMVAAAPVAVAAAAAP----- 520
         ||                    :|
Db   405 PNGSLDSVTYPVSPPPTSGPAAPPPPPPPPPPPPLPPPPPLPPLASLSHCGSQASPPP 464

Qy   521 ATAAATAAAVSPAAAGQIPAAASVASAAAVAP------ 573
         ||||||||||||||||||||||||||||||||     :|
Db   465 GTPLASTPSSKPSV---LPSPSAGAPASAETPLNPELGDSSASEPGLQAA-SQPAESPTP 520

Qy   574 -ALV---PVPAPAAAQASAPAQTOAPTSAPAVAPTPAPTPTPAVAQAEVPASPATPGPGH 629
         ||     : |  || |          ||  | | | : |  : |  |||: |||: |
Db   521 QGLVLGPPAPPPPPLPSGPAY-----ASALPPPGPPPPPL------ 566

Qy   630 RLSIPSLTCNPDKTDGP------ 659
         :|          :||       ::|
Db   567 PPPPPL---PNQAPPPPPPAPPLPASGIFSGSTSEDNRPL 606
```

```
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FH   region          1480..1659
FH   region          /label= internal_repeat
FH   region          1660..1839
FH   region          /label= internal_repeat
```

```
PN   WO9511309-A2.
PD   27-APR-1995.
PF   24-OCT-1994; U12162.
PF   22-OCT-1993; US-141239.
PR   22-OCT-1993; US-141239.
PA   (TEXA ) UNIV TEXAS SYSTEM.
PR   Lee W, Zhu X;
PI   Lee W, Zhu X;
DR   WPI; 95-170229/22.
DR   N-PSDB; Q86851.
PT   Purified mammalian protein mitosin and agents that bind it and
PT   inhibit its action - used to promote cell growth or to inhibit cell
PS   division and/or proliferation
CC   Claim 4; Fig 8B; 61pp; English.
CC   R72829 is human mitosin. Mitosin is involved in the regulation of
CC   the mammalian mitotic cell cycle. Mitosin as with E2F-1 (see R72824)
CC   interacts with the retinoblastoma protein (the retinoblastoma tumour
CC   suppressor gene product). Mitosin is first synthesised at the G1/S
CC   boundary, it is then phosphorylated from S through M phase, and during
CC   mitosis, is closely associated with the centromeres/kinetochores at the
CC   mitotic spindle poles. Mitosin is necessary for a eukaryotic cell to
CC   enter the M phase of the mitotic cell cycle and its degradation is
CC   necessary for a cell to advance on to the next stage. Mitosin is thus
CC   useful for controlling cell growth as overexpression of mitosin prevents
CC   a cell from exiting the M phase.
CC   An anti-mitosin antibody, antibody fragment or a phosphorylated mitosin
CC   mutein ( or nucleic acid encoding it) can also be used to inhibit cell
CC   division which is particularly useful for the study of the cell cycle.
CC   A further use is to control hyperproliferative cells, and so control
CC   diseases such as psoriasis and breast cancer. It can also be used to
CC   block gametogenesis of an immature gamete.
SQ   Sequence   2482 AA;
```

```
Query Match             6.8%;   Score 226.5;   DB 1;   Length 2482;
Best Local Similarity   21.4%;  Pred. No. 1e-05;
Matches 139; Conservative 100; Mismatches 233;   Indels 179;   Gaps 26;

Qy    26 QQMVELLSDENRLRQELE------                    -GCYEKVARLQKVETEIQRVSEAYENLVKSSSKRE 78
         ::|| |||  ||   || |
Db  1948 QDTLEVLQSSYKNLENELELTKMDKMSFVEKVNKMTAKEFELQR--EMHEMAQKTAELQE 2005

Qy    79 ALEKAMRNKLEGEIRRMHD-----                    -FNRDLRERLETANKQLAEKEYEGSED- 124
         ||: |||||||| |||  |           ::|||||||||||||||||||||||||
Db  2006 EL-SGEKNRLAGELQLLLEEIKSSKRDQLKELTLENSELKKSLDCMHKDQVEKEGKVREEI 2064

Qy   124 -----                    -TRKTISQLFAKNKESQREKEKLEAELATARSTNED)RRHIEI---RDQAL 170
         :| :|||||||||||||||||||||||||   ||    :||||
Db  2065 -AEYQLRLHEAEKKHQALLLDTNKQYEVEIQTYREKLTSKEECLSSQKLEIDLLKSSKEEL 2124

Qy   171 SNAQAKVVKLEEELKKKQV----                    -YVDKV-----EKMQQALVQLQAACEKREQLEHRLRTRL 222
         ||| || ||||||| |||    ||   ||||||||||||||||||||||||
Db  2125 NNSLKATTQILEELKTKMDNLKYVNQLKKENERAQGKMKLLIKSCKQLEEKEIIQKEL 2184

Qy   223 ERELESLRIQORGNCQPTNVSEYNA--AALMELLREKEERILALEADMTKWEQKYLEEN 280
         ||||||||||||||||||||||||   |||||||||||||||||||||||||||||||
Db  2185 S-QLQAAOEKQRGTGTVMDFKVDELITEIKELKETLEEK--                    -TKEADEYLDKY 2232

Qy   281 VMRHFALDAAATVAAQRDTTVISHSPNTSYDTALEARIQKEEEEILMANKRCLDMEGRIK 340
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2233 C---                    ------SLLISHE-                    ------KLEKAKEM------                    -LETQVA 2254

Qy   341 TLHAQIIEKDA----MIKVLQQRSRKEPSKTEQLSCMRPAKSLMSISNAGSGLLSHSSTL 396
         :|||||||||     ||||||||||||||||||||||||||||||||||||||||| |
Db  2255 HLCSQOSKQDSRGSPLLGPVVPGPSPIPSVTE------                    -KRLSSGONKASGKRQRSSGI 2306

Qy   397 TGSPIMEEKRDDKSWKGSLGILLLGGDYRAEYVPSTPSPVPPSTPLLSAHSKTGSRDCSTQ 456
         :|                                     :|::|
Db  2307 -----                    ------WENG----                    -SG------                    -PTPAPESFSKKSKKAVMSGIHPAE 2337
```

```
RESULT   5

Qy    457 TERGTESNKTAAVAPISVPAPVAAAATAAAITATAATI--TTTMYAAAPVAVAAAAPAA 514
              |||                                  |  |||
Db   2338 DTEGTE------FEPEGLPEVVKGFADIPTGKTSPYILRRTTM-------------- 2376

Qy    515 AAAPSPATAAAATAAAVSPAAAGQIPAAASVASAAAAVAPSAAAAAVQVAPAAPAPVPAPA 574
              |:||                          ||||||||||
Db   2376 ATRTSP-RLAAQKLAISPLSLGKENLAES-----SKPTAGGSRSQKVKVAQORSPVDSGT 2428

Qy    575 LVPVPAPAAAQOASAPAQTOAPTSAPAVAPPTPAPTPPAVAQOAEVPASPATG 625
              :  |||           ||  ||||||    :| :|||||
Db   2429 ILREP-----TTKSVPV-----NNLPERSPTDSPRECLRVKRGRLVPSPKAG 2470

RESULT   5

ID    W23996
AC    W23996;
DT    28-MAY-1998 (first entry)
DE    Human mitosin amino acid sequence.
KW    itosin; phosphoprotein; mitotic cell cycle; antibody; analogue;
KW    inhibition; M phase; Antagonist; hyperproliferative cell; cancer;
KW    leukaemia; lymphoma; chromosome segregation.
OS    Homo sapiens.
FH    Key             Location/Qualifiers
FT    Domain          258..280
FT                    /note= "leucine heptad repeat"
FT    Domain          340..362
FT    Domain          564..593
FT    Domain          1387..1443
FT    Domain          1885..1962
FT    Domain          2146..2188
FT    Domain          2165..2187
FT                    /note= "leucine heptad repeat"
FT    Misc_difference 2188
FT    Misc_difference 2300
FT                    /label= "Bipartite targeting motif"
FT    Misc_difference 2189
FT    Misc_difference 2301
FT    Misc_difference 2303
FT                    /label= "Optionally C or G"
FT                    /note= "optionally A or T"
```

W23996 standard; Protein; 2482 AA.

FT   domain   /label= Extended_coiled_structure
FT   1850. .2990
FT   domain   /label= Extended_coiled_structure
FT   3048. .3248
FT   /label= C-terminal_domain
FT   /note= "the C-terminal domain is predicted to
FT   form a proline-rich (10.6%) highly
FT   basic (pI 10) globular domain"

PN   WO9617867-A1.
PD   13-JUN-1996.
PF   08-DEC-1995; U16216.
PR   09-DEC-1994; US-353700.
PF   (FOXC-) FOX CHASE CANCER CENT.
PA   (UYTE-) UNIV TECHNOLOGIES INT INC.
PI   Rattner JB, Yen TJ;
DR   WPI; 96-287116/29.
DR   N-PSDB; T34578.

CC - DNA encoding kinetochore protein - used as a marker for the G2 and M
CC   phases of a cell cycle, partic. for detection of malignant diseases
PT   Diagnosing kinetochore protein, partic. for detection of malignant diseases
PN   claim 12; Page 41-54; 72pp; English.
PS   A 372 kDa human kinetochore protein, CENP-F (R99795), is detected
PS   by immunofluorescence microscopy only during the G2 and M phases
CC   of a cell cycle. It is the product of a cDNA clone (T34578) and
CC   isolated from a breast carcinoma cDNA library. Recombinant CENP-F
CC   can be produced by expression in prokaryotic or eukaryotic host
CC   cells. CENP-F can be used to detect autoimmune antibodies to
CC   the protein, which may provide an early diagnosis for the onset
CC   of various malignant diseases. Use of CENP-F as a cell cycle
CC   marker allows the specific detection of G2 and M phase cells.
SQ   Sequence 3248 AA;

Query Match   6.7%; Score 225; DB 1; Length 3248;
Best Local Similarity 21.4%; Pred. No. 1.8e-05;
Matches 110; Conservative 106; Mismatches 167; Indels 132; Gaps 21;

Qy   20 AIVSRAQQMVEILSDENRNLRQELEGGCYEKVARLQK----------------VETEIQRVSEAYE 68
           ::: | | . || .| | ..|| || ||| ||| | |
Db   2103 AEVKEKTELLQTLSSDVSELLKDKTHLQEKLQSLEKSQALSLTKCELENQIAQLNEKE 2162

Qy   69 NLVKSSSK-------------------REALEKAMRNKLEGEIR-----RMHDFNR---DLRE 104
           |||| :|| | | .:| | ||: | :| ||
Db   2163 LLVKESESLQARLSESDYEKLNVSKALEAALVEKGEFALRLSSTQEEVHQLRGIEKLRV 2222

Qy   105 RLETANKQ---LAEK--------EYEGSEDTRKTISQLFAKNKESQR---EKEKLEAFLAT 151
           |:| || | | || || | | | .| | ||| |||| .| |
Db   2223 RIEADEKKQLHIAEKLKEREENDSLKDKVENLERELQMSEENQELVILDAENSKAEVET 2282

Qy   152 ARSTNEDQRRIEIRDQALSNAQAKVVKLEEELKKKQVYVDKVEKM--------QQA 200
           |: : || |: | | || | .| | .| ||| .| ||||| | :|
Db   2283 LKTQIEMARSLKIFELDLVTLRSEKENLTKQIOEKQGOLSELDKLLSSFKSLLEEKEQA 2342

Qy   201 LVQLQ-------------------AACEKRE--------OLEHRLRTRLERE 225
           | || | : || .| || | .|
Db   2343 EIQIKEESKTAVEMLQNQLKELNEAVAALCGQOEIMKATEQSLDPPIEEEHQLRNSIEKL 2402

Qy   226 LESLRIQORQGNC--QPTNVSEYNAAAL----------MELLREKEERILALEADMTKWE 273
           |: : .. | | : |||| ||| || ||| || |
Db   2403 RARLEADEKQOLCVLQQLKESEHHADLLKGRVENLERELETARTNQEH-AALEAENSKGE 2461

Qy   274 QKYLEENV------MRHFALDAAA-------TVAAQRDTIVISHSP--NTYSYDTALEAR 317
           : | | ||| | | : | | || || | ||||||||
Db   2462 VETLKAKIEGMTQSLRGLELDVVTIRSEKENLTNELQKEQERISLEIINSSFENTLQ-- 2520

Qy   318 IQKEEEEILMANKRCLDMEGRIKTLHAQIIEKDAMIKVLQORSR-KEPSKTEQLSC--MR 374
           | | | | || | | | | |: | | | | | .|| || | : ||
Db   2520 -EKEQEKVQMKEKSSTAME-MLQTQLKELNERVAALHNDQEACKAREQNLSSQVELELE 2577

Qy   375 PAKSLMSISNAGSGLLSHSSTLTGSPIMEEKRDDK 409
           || : |: | || | || : || || ||||||
Db   2578 KAQLLQGLDEAKNNYIVLQSSVKG--LIQEVEDGK 2610

W26540
ID   W26540 standard; Protein; 442 AA.
XX
AC   W26540;
XX
DT   12-JAN-1998 (first entry)
XX
DE   Trypanosoma cruzi antigen TcH129.
XX
KW   Antigen; epitope; vaccine; protective immunity; Chagas disease;
KW   diagnosis; therapy; immunoassay.
OS   Trypanosoma cruzi.
OS   Trypanosoma cruzi Tulahen strain C2.
XX
PN   WO9718475-A1.
PD   22-MAY-1997.
PF   14-NOV-1996; U18624.
PR   14-NOV-1995; US-557309.
PF   (CORI-) CORIXA CORP.
PA   Houghton RL, Lodes MJ, Reed SG, Skeiky YAW;
DR   WPI; 97-289413/26.
DR   N-PSDB; T69165.

CC - Diagnosing Trypanosoma cruzi infection by detecting antibodies to
CC   novel antigens - which are useful in vaccines to provide protective
PT   immunity against Chagas' disease
PS   Disclosure; Page 80-82; 110pp; English.
PS   This polypeptide sequence comprises full-length antigen TcH129 of
CC   Trypanosoma cruzi, identified by sequencing a DNA clone (see T69165)
CC   obtained by screening a Trypanosoma cruzi genomic expression library
CC   with pools of sera from infected individuals. T. cruzi antigens
CC   (see W26530-41), or epitope-containing repeat sequences (see W19094-
CC   102, W19079-86 and W26542-44) of native antigens, can be used in a
CC   variety of immunoassays for detecting T. cruzi infection in a
CC   blood, serum, plasma, saliva, cerebrospinal fluid or urine sample.
CC   The polypeptides are also useful in vaccines and pharmaceutical
CC   compositions for inducing protective immunity against Chagas
CC   disease. They can be produced by expression in transformed or
CC   transfected host cells. TcH129 is a polymorph of the T. cruzi
CC   TcE epitope repeat sequence (see W19063).
SQ   Sequence 442 AA;

Query Match   6.5%; Score 216; DB 1; Length 442;
Best Local Similarity 44.9%; Pred. No. 5e-06;
Matches 71; Conservative 7; Mismatches 58; Indels 12; Gaps 6;

Qy   476 APVAAAAAITATAAAITITTMVAAAPVAVAAAAAPAAAAAPSPATAAAATAAAAVSPAAA 535
           | || | |||| ||| | | || |||| || :||| |
Db   281 AAAAAKQKAAAAKAAAPSGKKSAKAAAP--AKAAAAPAKAAAPPAKTAAAAPAKAAAPKA 338

Qy   536 GQIPAAASVASAAAVAPSAAAAAAVQVAPAAPAPV---PAPALVPPAPAAA----QASA 588
           | | | | | | | | ||| | | | ||
Db   339 AAPPAKAAAPPAKTAAPPAKTAAPPAKAAAAPPAKAAAAPPAKAAAAP-PAKAAAAPAKAAAA 397

Qy   589 PAQTQAPTSAPAVAPTTPAPTTPTPAVAQAEVPASPATGP 626
           | :|| | | || || || || | | || |
Db   398 PAKAAAPPAKAAAPPAKAAAP-PAKAAAP-PAKA-AAP 433

W00024
ID   W00024 standard; Protein; 1972 AA.
XX
AC   W00024;
XX
DT   25-MAR-1997 (first entry)
XX
DE   Smooth muscle myosin heavy chain SM1 isoform protein.
XX
KW   smooth muscle; myosin heavy chain; SM1 isoform; rabbit; arteriosclerosis;
KW   gene therapy; mouse; myosin heavy chain; SM2 isoform; retrovirus; adenovirus; restenosis;
KW   associated adenovirus; coronary artery catheterisation; sclerotic artery.
OS   Mus musculus.
XX
PN   WO9623065-A1.
PD   01-AUG-1996.
PF   25-JAN-1996; J00134.
PR   25-JAN-1995; JP-010085.
PF   (OSAP) OSAKA PREFECTURAL GOVERNMENT.
PA   (VESS-) VESSEL RES LAB CO LTD.
PA   Arakawa E, Hasegawa K, Ishiyama H, Matsuda Y, Oda S;
PA   Shigawara M, Takahashi K;
DR   WPI; 96-362093/36.
DR

DR   N-PSDB; T34291.
PT   DNA encoding smooth muscle myosin heavy chain SM1 isoform - used in
PT   a vector for gene therapy for reduction of re-stenosis following
DR   coronary artery catheterisation
DR   N-PSDB; T46148.
PS   Claim 5; Page 14-27; 42pp; Japanese.
PT   This sequence represents the smooth muscle myosin heavy chain SM1 isoform
PT   protein. The SM1 coding sequence was isolated from a mouse embryo cDNA
CC   library using probes based on the rabbit smooth muscle myosin heavy chain
CC   protein. The isolated coding sequence is included in the gene
CC   therapy vector of the invention. The vector of the invention consists of
CC   the coding sequence inserted into a retrovirus, adenovirus, associated
CC   adenovirus or animal-expression plasmid vector (such as pCXN2 or
CC   PAGE208). The vector can be used in the gene therapy treatment of
CC   arteriosclerosis, particularly for the reduction of restenosis occurring
CC   following coronary artery catheterisation for widening of sclerotic
CC   arteries.
SQ   Sequence   1972 AA;

Query Match   6.4%; Score 213; DB 1; Length 1972;
ches   86; Conservative   85; Mismatches 103; Indels 122; Gaps 14;

Qy   30 EILSDENRNLR------------QELEGCYEKAVARLQKVEIE------IQRVSEAYENLVKSSSKREA   79
            ::|:::|:| ::|| |||: | | :   : ||: :|:: |: | |
Db   1554 DVQATEDAKRLRLEVNMQALKGQOFERDLQARDEQNEEKRRQLQROLHEYEIELEDERKQRA 1613

Qy   80 LEKAMRNKLEGEIRRMHDFNRDLRERLETFANKQLAE-------KEYEGS-EDTRKYISQLFA 133
            | | ||||||||||| : ||| ||| |||| : ::|| :: |
Db   1614 LAAAAKKLEGDLKDLELQADSAIKGREEAIKQLRKLQAQMKDFQRELDDARASRDEIFA 1673

Qy   134 KNKESQR------------------EKEKLEAELATA----RSTNEDQRRH 162
            ::|::| |||||:|||| :|:| ||||
Db   1674 TSKENEKKAKSLEADLMQLQEDLAAAERARKQADLEKEELAASSLSGRNTLQDEKRR 1733

Qy   163 IEIRDQALSNAQAKVVKLEEEIKRKQVYVDKV-EKMQQALVQLQAACEKREQLEHRLRTR 221
            ::|||||:|||||:||||||||||||||||| |::||||||||||||||||||:|| |
Db   1734 LE----------ARIAQLEELEEEQGNMEAMSDRVRKATLQ---------AEQLSNELAT- 1776

Qy   222 LERELESLRIQQROGNCQPPTNVSEYNAAALMELLREKERILALEADMTKWEQKYLEENV 281
            : |:|:| |:|:: || |: || || :||||||| | :| ||||:|:||||
Db   1776 -ER---------STAQKNESARQQLERQNKE--------------------------- 1797

Qy   282 MRHFALDAAATVAAQRDTTVISHSPNTSVDTALEARIQKEEEEI--------LMANKRCLD 334
            :| | | :|::|||| |: | : | : ||:||||||||||||| | | |||
Db   1797 LRSKLQEVEGAVARKLKSTV------------------AALEAKIAQLEEQVEQEAREKQAATKSLKQ 1846

Qy   335 MEGRIKTLHAQIIEKDAMIKVLQQRSRKEPSKTEQL 370
            ::|::| ||||||| : ||:|:|| |: :|| ||
Db   1797 AAPAKAATAPAKAAT-APAKA-AAAPAKAATAP 1882

Db   347 KDKKLKEVLLQVEDERKMAEQYKEQAEKGNTKVKQL 1882

RESULT   9
W06913
ID   W06913 standard; Protein; 262 AA.
AC   W06913;
DT   02-APR-1997 (first entry)
DE   T. cruzi L19E homologue TcE.
KW   L19E homologue; TcE; diagnosis; infection; antigen; Chagas disease;
KW   leishmaniasis.
OS   Trypanosoma cruzi strain MHOM/CH/00/Tulahuen.
FH   Key   Location/Qualifiers
FT   region   1..143
FT   /note= "polypeptides comprising amino acids 1-143
FT   are useful in methods for screeninf for T.
FT   cruzi and Leishmania infection"
FT   region   137..247
FT   /note= "amino acids 137-247 comprise 16 copies of
FT   a tandemly arrayed 7-amino acid repeat"

PA   (CORI-) CORIXA CORP.
PI   Reed SG;
WPI; 96-485445/48.
DR   N-PSDB; T46148.
PT   Claim 1; Fig 2; 59pp; English.
PT   Detecting and preventing T. cruzi infection - using polypeptide(s)
PT   or antibodies contg. or reactive with antigen epitope(s) of T. cruzi
PT   proteins.
CC   The 35 kDa T. cruzi TcE antigen (W06913) shows homology in its
CC   N-terminal portion to the eukaryotic ribosomal protein L19E. The
CC   C-terminal portion comprises a 7-amino acid repeat region. The
CC   TcE amino acid sequence was deduced from a cDNA clone (T46148). The
CC   isolated from an amastigote library. Epitopes such as TcEr
CC   (W06914) derived from the TcE repeat region can be used in methods
CC   for detecting T. cruzi infection (Chagas' disease) and in vaccine
CC   compns. Polypeptides comprising the L19E homologous region can be
CC   used in methods for detecting Leishmania or T. cruzi infection.
SQ   Sequence   262 AA;

Query Match   6.4%; Score 213; DB 1; Length 262;
Best Local Similarity   48.4%; Pred. No. 3.9e-06;
Matches   74; Conservative   9; Mismatches 54; Indels 16; Gaps 9;

Qy   479 AAAATAAAATTATAATTTITTTWWAAAP---VAVAAAAAPA-AAAAPSPATAAAAVSPAA 534
            |||| || |||||| | || ||| || ||| |||| ||: |||
Db   112 AAAAAAAAKQKAAAKK-----AAAPSGKKSAKAAIAPAKAAAAPAKAAAAPAKAAAAPAK 165

Qy   535 AGQIPAAAASVASA-AAVAPSAAAAAAAVQVAPAPAPAPALIVPVPAPAAAAQASAPAQTQ 593
            | | | :| | || |||| | |: || || || || |||||| |||: |||
Db   166 AAAAPAKAAAAPAKAAAATAPAKAAAAPAKTA-AAPAKAAAPAKA-AAAAPAKA-ATAPAKAA 222

Qy   594 APTSAPAVAPTPAPTPTPAVAQAEVPASPATGP 626
            | ||| |:| ||| : || ||||||||||||
Db   223 AAPAKAATAPAKAAT-APAKA-AAAPAKAATAP 253

RESULT   10
W89253
ID   W89253 standard; Protein; 1274 AA.
AC   W89253;
DT   10-MAR-1999 (first entry)
DE   Human ALP.
KW   PTP04; PTP05; PTP10; SAD; ALP; ALK-7; protein tyrosine phosphatase;
KW   type I receptor serine/threonine kinase; cancer; leukaemia; lymphoma;
KW   neurodegenerative disease; neuronal survival; Alzheimer's disease;
KW   Parkinson's disease; Huntington's disease.
OS   Homo sapiens.
PN   WO9849317-A2.
PD   05-NOV-1998;
PF   27-APR-1998;
PR   23-OCT-1997; U08439.
PR   28-APR-1997; US-063595.
PR   20-MAY-1997; US-044428.
PR   11-JUN-1997; US-047222.
PR   11-JUN-1997; US-049477.
PR   18-JUN-1997; US-049756.
PA   (SUGE-) SUGEN INC.
PI   App H, Clary D; Courteidge SA, Hui TH, Jallal B,
PI   Markby D, Onrust S; Peles E, Plowman GD;
DR   N-PSDB; V81748.
WPI; 99-009434/01.
PT   Claim 2; Page 160-164; 193pp; English.
PT   New nucleic acid encoding specific protein tyrosine phosphatases -
PT   useful for identifying specific modulators for treatment and
PT   prevention of cancer and neurodegenerative disease
CC   The present invention describes isolated, enriched or purified nucleic
CC   acids encoding PTP04, SAD, PTP05, PTP10, ALP and ALK-7 proteins. The
CC   present sequence represents human ALP. The above proteins, other than
CC   ALK-7, are protein tyrosine phosphatases (PTPs) and are used to identify
CC   substances that modulate their activity (i.e. agonists and antagonists,
CC   including NBP) in vivo or in vitro. These substances are used to treat
CC   or prevent diseases associated with abnormal signal transduction

RESULT 11
W02258
ID   W02258 standard; Protein; 1411 AA.
AC   W02258;

CC   pathways that involve the proteins, particularly cancer (e.g. leukaemia
CC   and lymphoma), while modulators of ALK-7 (which is a type I receptor
CC   serine/threonine kinase) are used to promote neuronal survival,
CC   particularly for treating Alzheimer's, Parkinson's or Huntington's
CC   diseases. Nucleic acid fragments of the polynucleotides encoding the
CC   proteins can be used as probes to identify and clone related sequences;
CC   to detect protein-encoded RNA; to generate transgenic animals and in
CC   gene therapy (optionally after mutation). Ab are used to determine the
CC   proteins.
SQ   Sequence   1274 AA;

Query Match   6.3%; Score 212.5; DB 1; Length 1274;
Best Local Similarity   20.7%; Pred. No. 3; 1e-05;
Matches   151; Conservative 104; Mismatches 306; Indels 167; Gaps 30;

Qy   25 AQOMVEILSDENRNLRQELEGCYE-KVARLQKVETEIQRVSEAVENL-VKSSSKREALEK 82
         :   : :   : ||||| || | | | | | | |
Db   3 AHEASSLYSEEKAKLLREMAKIEDKNEVLDQFMDSMQLDPETVDNLDAYSHIPQLMEK 62

Qy   83 --------AMRN----KLEGEIRRMHDFNRDLRERLETANKQLAEKEYEGSEDTRKT 127
                        :|| | | : : | ||| :|: || || |
Db   63 CAALSVRPDTVRNLVQSMQVLSGVETVDVEASLKDIRDLLE--EDELLEQKFQ------ 114

Qy 128 ISQLFAKNRESQREKEKLEAELA----------TARSTNEDQRRHIEIR--------DQA 169
         :|| || | :| || | | : | | : |||| ||| |
Db 115 VGQAGAISITSKAELAEVRREWAKYMEVHEKASFTNSELHRAMNLHVGNLRLLSGPLDQV 174

Qy 170 LSNAQAKVKLEEELKKKQVYVDDKVVEKMQQALVQLQAACEKREQLEHRLRTRLERELESL 229
         || | | |||||| | || : || | ||| ||||||||| |||||| ||
Db 175 RAALPTPALSPEDRA-----VLQNLKRILAKVQEMRDQRVSLEQQLRELIQKD---- 223

Qy 230 RIQORQRGNCQOPTNVSEYNAAALMELLREKEERILALEADMTKWEQ--KYLEENVMRHFAL 287
                :|| ::| : | :|| : | | | | | | ||: ||
Db 223 --------DITASLVTTDHSEMKKL-FEEQLKKYDQLKYVLEQNLAAQDRV 264

Qy 288 DAAAT-----------VAAQRDTTVISHSPNTSYDT-------ALEARIQKEEEILMANKRCLDM 335
                                  | | :   :   :|| ::| :   :||| |||||
Db 265 LCALTEANVQYAAVRRVLSDLDQKWNSTLQTLVASYEAYEDLMKKSQE----GRDFYADL 320

Qy 336 EGRIKTL---------HAQIIERDAMIKVLQQRSRKEPSKTEQLSCMRPAKSLMS 381
         || | | | |: | ||||| | | : || | |
Db 321 ESKVAALLERTQSTCQAREAARQOLLDRE------LKKKPPRPTAPKPLLPRREESEAVE 375

Qy 382 ISNAGSGLLSHSSTLTGSPIMEEKRDDKSWKGSLGILLGGDYRAEYVPSTPSPVPPSTPL 441
             | : | | | | | : | : | : || | || ||||||
Db 376 AGDPPEELRSLPPDMVAGPRLPD------TFLGS-------ATPLHFPSPFPSST-- 419

Qy 442 LSAHSKTGSRDCSTOTERGTESNKTAAVAPISVPAPVA-AAATAAAITATAARTITTTMV 499
         || | | | : | : | || | | | |||||||||||| :| : |
Db 419 IQAGEGETAVLNQLQEKNHTLQEQVTQLFEKLKNQSESHKQAQENLHDQVQEQKAHLRAA 472

Qy 500 AAAPVAVAAAAAAPAAAAAPSPATAAAATAAAAVSPAAAGQI----PAAASVASAAAVAPSA 554
             | | | :: | : ||||| || || |||||
Db 473 PRSSPQHGVVSSPYVGVGPAPVVAGLPSAPPQOFSGPELAMAVRPATTVDSIQAPIPSH 532

Qy 555 AA-------AAAVQVAPAAPAPVPAPALVPVPAPAAAQASAPAQ---TQAPTS--APAVAPT 604
         || | :: ||| | | ||| | || | ||||| ||||||
Db 533 TAPRPNPTPAPPPCFVPVPPQPLPTPYTYPAGAKQPIPAQHHFSSGIPTGPPAPRIGPQ 592

Qy 605 PAPTPTPAVAQAEVPASPATGCPHRLSIPSLTCNPDKTDGPVFHSNTLERKTPIQ---- 661
         || | || | | || | || | | | :: : | :| | ||
Db 593 FQPHPQPHPSQAFGPQPPQO---PLPLQHPHLF--PPQAPG----LLPPQSPYPYAPQ 641

Qy 661 --ILGQEP 666
           :||||
Db 642 PGVLGQPP 649

DT   09-MAR-1997 (first entry)
DE   Nucleolar/endosomal auto-antigen p162.
K9   Auto-antibody; p162; rheumatic disease; antigen; diagnosis;
KW   gene therapy.
GS   Homo sapiens.
PN   DE19515514-C1.
PD   12-SEP-1996.
PF   27-APR-1995; 015514.
PR   27-APR-1995; DE-015514.
PA   (PRIV) PRIVATES INST IMMUNOLOGIE & MOLEKULARGEN.
PA   Renz M, Seelig HP;
DR   WPI; 96-401153/41.
DR   N-PSDB; T58751.
PT   DNA encoding nucleolar-endosomal auto-antigen - useful for exact
PT   diagnosis of rheumatic disease, in gene therapy and for removal of
PS   specific auto-antibodies
PS   Claim 1; Fig 2; 15pp; German.
CC   Transformed cells can be cultured to produce the antigen p162, for use
CC   in exact (differential) diagnosis of rheumatic disease, i.e. they
CC   can detect, in immunoassays, Western blots, etc., rheumatism-
CC   specific auto-antibodies. The antigen can be used therapeutically,
CC   in the removal of auto-antibodies from the circulation, or when
CC   coupled to a cytotoxin, the elimination of auto-antibody-
CC   producing lymphocytes.
SQ   Sequence   1411 AA;

Query Match   6.3%; Score 211; DB 1; Length 1411;
Best Local Similarity   22.2%; Pred. No. 4; 4e-05;
Matches   108; Conservative 75; Mismatches 155; Indels 148; Gaps 17;

Qy   30 EILSDEN--RNLRQELEGCYEKVARLQKVSTEIQRVSEAVENLVKSSSKREALEKAMRNK 87
         :|||||| ||||||||| ||||||||| ||||||||||||||||||||||||||
Db 228 FSVSKKNIQATLHQKDLDCQOLQSRLSASETSLHRIH-------VELSEKGEATQKLKEEL 381

Qy   83 LEGEIRRMHDFNRDLRERLETANKQLAEKEYEGSEDTRKTISQLFAKNRESQREKEKLEA 147
         ||||||||||||||||||||||||||||||||||||||| |||||||||||||||||||
Db 382 SEVEFKYQH-----IKAEFKQLQOQREEKEQHGLQ-LQSEINQLHSKLLETERQLGEAHG 435

Qy 198 QOALVQLQAA--CEKREQLEHRLR------------TRLERELESL--R 230
         ||||||||| |||||||||| ||||||||| |
Db 436 RLKEQROLSSEK---LMDKEQOVADLQLKLSRLEEQLKEKVTNSTELQHQLDKTKQOHQE 492

Qy 231 IQQRQRGNC---------QPT-----NVSEYNAAALMEL------LREK 258
         |||||||| || |||||||||| ||||
Db 493 QOALQOSTTAKLREAQNDLEQVLRQIGDRKDQKIQNLEALLQKSKENISLLEKEREDLYAK 552

Qy 259 EERILALEADMT-----KWEQKYLEENVMRHFALDAAATVAQR------- 298
         |||||||||| ||||||||||||||||||||||||||
Db 553 IQAGEGETAVLNQLQEKNHTLQEQVTQLFEKLKNQSESHKQAQENLHDQVQEQKAHLRAA 612

Qy 298 --------------------DTTVISHSPNTSYDTALEARIQKEEEILMANKRCLDMEGRI 339
                              ||||||||||||||||||||||||||||||||||||||
Db 673 QNALQDFQOELNKITTOLDQVTAKLQDKOEHCSQLESHLKEYKEKYLSLEQKTEELEGQI 732

Qy 340 KTLHAQIE----KDAMIKVLQQRSR----------KEPSKTEQLSCMRPA 376
         |||||| ||||||||||||| |||||||||||||
Db 733 KKLEACSLEVKASKEQALQDLQOQRQLNTDLERATELSKQLEMEKEIVSSTRLDLQKKS 792

Qy 377 KSLMSI 382
         |||||
Db 793 EALES! 798

```
DE   Human mDia Rho targeting protein.
KW   Rho Protein; mDia; mammalian diaphanous; target protein; human;
KW   Rho protein-combining; proline-combining; chromosome 5q31.2; disease;
KW   respiratory tract; oversensitiveness; bronchial asthma; marrow leukaemia;
KW   marrow dysplasia syndrome.
OS   Homo sapiens.
PN   J1026268-A.
PD   06-OCT-1998.
PF   25-MAR-1997; 090170.
PR   25-MAR-1997; JP-090170.
PA   (KIRI ) KIRIN BREWERY KK.
DR   WPI; 99-002481/01.
DR   N-PSDB; V62993.
SQ   Sequence  1315 AA;
CC   dysplasia syndrome.
CC   oversensitiveness, bronchial asthma, acute marrow leukaemia and marrow
CC   for the elucidation of mechanism of diseases such as respiratory tract
CC   used for the recombinant production of the protein. The protein is useful
CC   gene is positioned at human chromosome 5q31.2. The nucleic acid can be
CC   dodecyl sulphate-polyacrylamide gel electrophoresis (SDS-PAGE) and its
CC   has proline-combining ability, has a M.W. of 150 kDa measured by sodium
CC   diaphanous). This protein has active type Rho protein-combining ability,
CC   This sequence represents a human Rho target protein, mDia (mammalian
PT   Claim 2a; Page 28-31; 54pp; Japanese.
PT   mechanisms of respiratory tract disease
PS   Human Rho target protein and its gene - useful for elucidation of

Query Match       6.3%; Score 209.5; DB 1; Length 1315;
Best Local Similarity   20.5%; Pred. No. 5e-05;
Matches 163; Conservative  97; Mismatches 280; Indels 257; Gaps  39;

Qy    7 SSASQPVP-ADPFAIVSRAQQMV---EILSDENRN----------LRQELEGCYEK   49
        |||||                    :| ::|            :||||||
Db  114 SSASYGDDPTAQSLQDVSDEQVLFEQMLLDMNLNEEKQQPLREKDIIKREMVSQYLY  173

Qy   50 VARLQKVETEIQRVSEAYENLVKSS----------SKREALEKAMRN--------K   87
        | |:                       ::|      | :||:|:  |          |
Db  174 TSKAGMSQKESSKSAMMYIQELRSGLRIDMPLLSCLESLRVSLNNNPVSWVQTFGAEGLAS  233

Qy   88 LEGEIRRMHDFNRDLRERLETANKQLAEKEYEGSEDIRK----TISQLFAKNKESQREK  142
        ::||                       ||             ||
Db  234 LLDILKRLHD-----EKEETA----------GSYDSRNKHEIIRCLKAFMNNKFG----  274

Qy  143 EKLEAELATARSTNEDQRRHIEIRDQALSNAQAKVVKL---------EEELKKK-----  188
                                                          :| |
Db  274 -----IKTMLETEEGILILLVRAMDPAVPNMMIDAAKLLSALCILPQPEDMNERVLEAMT  327

Qy  188 -QYVYDKVEKMQQAL---------VOLOAAC-----EKREQLEHRLRTRLE--------  224
        :| ::|| |                :|:|:                :| ||:
Db  328 ERAEMDEVERFQPLLDGLKSGTTIALKVGCLQLINALITPAEELDFRVHIRSELMRLGLH  387

Qy  224 -----RELES--LRIQ---QROQNCQP------TNVSEYNAA--ALMELLREK  258
             :||:    ::|                    :| :         ||:
Db  388 QVLQDLREIENEDMRVQLNVFDEQGEEDSYDLKGRLDDIRMEMDDFNEVFQILLNTVRDS  447

Qy  259 E------------ERIALEADMTKWDQKY--LEENV--------MRHFALDAAAT  292
        |            ::|                ::
Db  448 RAEPHFLSILQHLLLVRNDYEARPQYYKLIEECISQIVLHKINGADPDFKCRHLQIEIGL  507

Qy  293 VAAQRDTTVISHSP------NPSYDTPALEARIQKEEEEILMANKRCLDMEGRIKTLHAQI  346
        :| |:            | |                | :|            ::|
Db  508 IDQMIDKTKVEKSEAKAAELEKKLDSELTARHELQVEMKKMES----DFEQKLQDLQQ--  562

Qy  347 IEKDAMIKVLQQRSREPESKTEQLSCMRPAKSLMSISNAGSGLLSHSSTLIG--SPIMEE  404
         ||:|                                                  ::
Db  562 -EKDAL-----HSEKQQIATEK----------QDLEAEVSOLIGEVAKLTKE  597

Qy  405 KRDDKSWKGSLGILLGGDYRAEVVPSTPS--PVPPSTPLLSAHSKTGSRDCSTQTERGTE  462
                                        |||||
Db  598 LEDAKKEMASLSAA-------AITVPPSVPSRAPVPPAPL------PGDSGT-  638

Qy  463 SNKTAAVAPISVPAPVAAAATAAAITAAAITITTMYAAAPVAVAAAAAPA----AAAA  517
```

```
RESULT 13
W63043
ID   W63043 standard; Protein; 561 AA.
AC   W63043;
DT   26-OCT-1998 (first entry)
DE   Streptococcus uberis bovine lactoferrin binding protein.
KW   Bovine lactoferrin binding protein; LBP; mastitis; vaccine;
KW   diagnosis.
OS   Streptococcus uberis strain su-1 (ATCC 9927).
FH   Key            Location/Qualifiers
FT   Peptide        1..51
FT                  /label= Sig_peptide
FT                  /note= "alternative translation start site at
FT                  Met-11"
FT   Protein        52..561
FT                  /label= Mat_protein.
FT   Region         148..199
FT                  /note= "central repeated amino acid sequence A1"
FT   Region         200..212
FT                  /note= "central repeated amino acid sequence B1"
FT   Region         213..271
FT                  /note= "central repeated amino acid sequence C1"
FT   Region         282..325
FT                  /note= "central repeated amino acid sequence A2"
FT   Region         326..339
FT                  /note= "central repeated amino acid sequence B2"
FT   Region         340..397
FT                  /note= "central repeated amino acid sequence C2"
FT   Peptide        525..530
FT                  /note= "surface anchor motif"
PN   WO9821231-A2.
PD   22-MAY-1998.
PF   14-NOV-1997;
PR   14-NOV-1996; CA0867.
PR   14-NOV-1996; US-031117.
PA   (UYSA-) UNIV SASKATCHEWAN.
PI   Jiang M, Maclachlan PR, Potter AA;
DR   WPI; 98-297860/26.
DR   N-PSDB; V42601.
PT   Immunogenic Streptococcus uberis protein(s) that bind bovine
PT   lactoferrin ; associated regulatory protein, useful in vaccines for
PT   treatment and prevention of mastitis
PS   Claim 2; Fig 2A-C; 105pp; English.
CC   This is the bovine lactoferrin binding protein (LBP) of
CC   Streptococcus uberis su-1. Its amino acid sequence was deduced
CC   from the novel isolated lbp gene (see V42601). The LBP is
CC   lactoferrin species-specific; human lactoferrin does not
CC   effectively block binding of bovine lactoferrin. The invention
CC   provides recombinant vectors, transformed host cells and methods of
CC   producing recombinant bovine LBP of S. uberis. The bovine LBP,
CC   immunogenic fragments and/or chimeric proteins can be used, either
CC   alone or in combination with other antigens, in novel subunit
CC   vaccines for the prevention and treatment of S. uberis infections,
CC   particularly mastitis, as well as in diagnostic methods for
CC   determining the presence of S. uberis infections.
SQ   Sequence  561 AA;

Db  638 ------------IIPPPAPGDST-----TPPPPPPPPPPPPGGVCISSPPSLPGGTAISP  684
                     |: :|                :| |
Qy  518 PSPAPAAAATAAAVSPAAAG-QIPAAAASVASAAAAVAPSAAAAAVQVAPAAPAPVPAPA-L  575

Db  685 PPPLSGDAITPPPPLPEGVGIPSPSSLPGGTAIPPPPPLPGSARI-PPPPPPLPGSAGI  743
Qy  576 VPVPAPAAAOASAPAQTQAPTSAPAVAPPP----------APTTPAVAQAE------VPAS  621

Db  744 PPPPPLPGEAGMPPPPPPLPGGPGIPPPPPGMGMPPPPPFGFGVPAA  803
Qy  622 PAT--GPGPHRLSIPSL  636

Db  804 PVLPFGLTPKKLIKPEV  820
```

RESULT 14

```
Query Match     6.28;    Score 207.5;  DB 1;  Length 561;
Best Local Similarity   22.1%;  Pred. No. 2.3e-05;
Matches 107;  Conservative  85;  Mismatches 157;  Indels 135;  Gaps  20;

Qy    53 LQKVETEIQRVSEAYENLVKSSSKREALEKAMRNKLEGEIRRMHDFNRDLRER---LETA 109
         |:|| :| :|:| |:|::|| ||: | ::|:||:|:|:||||||||||||||   |:|
Db   148 VQEVRSRLNEKDELDEEL--SNKKEELQK-LTEKIEKTIKEKENLNKEITEKNSEISKM 203

Qy   110 NKQLAEKEYEGSE-----------DTRKTISQLFAKNKE----SQREKEKLEAEL 149
         :|||||||||||             ||||| :|| :|      ::|:|||||||
Db   204 EEELSREKEIAENKEELADALGELFDAEETIDKKFARVKDLTEKLDASRKEHEALAKEF 263

Qy   150 ATARSTNEDQRRHIEIRDQALSNAQAKVKLE----EELKKQVYVD-------KVEKMQ 198
         |||||:| |||:|:||||||||||||||||     |||||||| |       |:||:|
Db   264 AESQKGYE---KELADKHTALGEAEKRNADLEAGNKELKENLEMAEGISDDLQKKVNKAE 320

Qy   199 QALVQLQAACEK-REQLEHRLRTRLERELESLRIQORGQNCQPTNVSEYNAAAL-MELLR 256
         |:||:||||: | :|:||:|:||||:| :  |:|:|:|||   ||: |:|| :| |:||
r    321 QEMKELSAQLEEAKEELETEKAKLAESEKENAKLTE------------ERDAAKREAEKVP 369

Qy   257 EKEERILALEADMT--KWEQKYLE----ENVMRHFALDAAATVAAQRDTIVISHSPNTSYD 311
         |||||||||||:    :|:||:|     |||||||||||||||:||||||||||||:||||
Db   370 ELEEQQVEKLVEEITAARKEAEELQAKAEGLEKDF------EAVKAEKE-----------  412

Qy   312 TALEARIQKEEEILMANKRCLDMEGRIKTLHAQIIEKDAMIKVLQ---QRSREPSKTE 368
         ||||||||||||:|| |: | |: :|:||| ||:|||||||||:    :| |:|| ||
Db   412 -ALEAEIAKLKE-----DHOKEVDALNALLADKEKMLKNLQDOLDKAKEEAMKNE 460

Qy   369 QLSCMRPAKSLMSISNAGSGLLSHSSTLTGSPIMEEKRDDKSWKGSLGILLGGDYRAEYV 428
         |||||||||: | | :| || || :| |||:||||||||||||: | | |||||||: |
Db   461 QMSQEEKAKLQAELDQAKKEL------------------------------AEKI 485

Qy   429 PSTPSPVPPSTPLLSAHSKTGSRDCSTOTERGTESNKTAAVAPISVPA------PVAAAATA 484
         |: || :||| || | |:|||:|:|:|| |:|:||||: | |: |:      || ||||
Db   486 KDMPNKVAP-----OAEGKANAQOAAPNQNQNNOAQANQAKNGNNLPSTGDKPVNPLLVA 540

Qy   485 AAIT 488
         :: |
Db   541 SGLS 544
```

RESULT 14

```
ID  W54241  standard; Protein; 1886 AA.
AC  W54241;
    17-AUG-1998  (first entry)
PN  Rattus norvegicus mutant alpha-myosin heavy chain.
KW  Actin binding domain; alpha myosin heavy chain; bridging; mouse;
KW  transgenic mammal; congestive heart failure; study; treatment; diet;
KW  exercise; effects; identification; hypertrophic cardio-myopathy;
KW  dilated or hypertrophic cardiomyopathy; acute aortic regurgitation;
KW  tricuspid stenosis; constrictive pericarditis; hypertension;
KW  acute infective endocarditis; ischaemic heart disease;
KW  primary myocardial disease; valvular disease; pericardial disease;
KW  hyperthyroidism; anaemia; arteriovenous fistula; beri-beri;
KW  Paget's disease; transgene.
OS  Rattus norvegicus.
OS  Synthetic.
FH  Key            Location/Qualifiers
FT  Region         403
FT                 /note= "Arg403Gln mutation"
FT  Region         468..475
FT                 /note= "nonmyosin actin binding domain bridge"
```

```
PN  WO9813476-A1.
PD  02-APR-1998.
PF  26-SEP-1997; U17296.
PR  26-SEP-1996; US-026742.
PA  (LEIN/) LEINWAND LA.
PI  Vikstrom KL;
WPI 98-230690/20.
DR  N-PSDB: V21518.
PT  Transgenic mouse models for congestive heart failure and
PT  hypertrophic cardio-myopathy - used to study molecular and cellular
```

```
PT  events, identify potential therapeutic agents, assess effects of
PT  diet etc.
PS  Disclosure; Pages: 53-58; 75pp; English.
CC  The sequence is that of a mutant rat alpha-myosin heavy chain
CC  which was used in the development of transgenic mammals,
CC  specifically mice. They can be used as a model for studying congestive
CC  heart failure (CHF) or hypertrophic cardio-myopathy. Such animals
CC  are used to study molecular and cellular events associated with
CC  CHF; to identify compounds for treating CHF, and in evaluating
CC  effects of diet and exercise on CHF. Conditions associated with CHF
CC  that can be evaluated this way are dilated or hypertrophic
CC  cardiomyopathy; acute aortic regurgitation; tricuspid stenosis;
CC  constrictive pericarditis; acute infective endocarditis; ischaemic
CC  heart disease; hypertension; primary myocardial disease; valvular
CC  or pericardial disease; hyperthyroidism; anaemia; arteriovenous
CC  fistula; beri-beri and Paget's disease.
SQ  Sequence  1886 AA;
```

RESULT 15

```
Query Match     6.18;    Score 205.5;  DB 1;  Length 1886;
Best Local Similarity   21.9%;  Pred. No. 0.00014;
Matches 116;  Conservative  84;  Mismatches 177;  Indels 153;  Gaps  22;

Qy    21 IVSRAQOMVEILSDENRNLRQLELEGCYEKVARLQKVETEI-QRVSEAYENLVKSSSKREA  79
         |:||||| ||:|::||: :|::|: ||:|:| :|| ||||: |:|| |||||:|||::||
Db    937 IIAKLTKEEKALQEAAHQQALDDLQAEEDKVNTLTKSKVRLEQOVDDLEGSLEQEKKVRMD  996

Qy    80 LEKAMRNKLEGEIRRMHDFNRDLRERLETANKQLAEKEYEGS------EDTRKTISQLFA 133
         || |:||:||||||||||||||||||| |||||||||||||       ||||||||||||
Db    997 LERAKR-KLEGDLKLTQESIMDLENDKLQLEEKLKKKEFDISQQNSKIEDEQALALQLQK 1055

Qy   134 KNKESQREKEKLEAEL---ATARSTNE---DQRRHIEIRDQALSNA-QAKVVKLEEELK 185
         :|||||||||||:| |   || |: | |  |:|:||:||||||| | |||: ||:|||
Db  1056 KLKENQARIELELEEIEAAERTRAKVEKLRSDLTRELEEISERLEEAGGATSVQIEMNKK 1115

Qy   186 KKQVYVLKVEKMQQALVQLQA------------ACEKREQLEH--RLRTRLERELSLRI 231
         ||| : |::  :||:|| |||            |:|||||| |  ||||||||||:|||
Db  1116 REAEFQKMRRDLEEATLQHEATAAALRKKHADSVAAELGEQIDNLQRYKQKLEKEKSEFKL 1175

Qy   232 Q------------------OROGNCQPTN-----VSEYNA 250
         :                  :| |: |||     ||||||
Db  1176 ELDTVTSHMEQIIKAKANLEKVSRTLEDQANEYRVKLEEAQRSLNDFTTQRAKLIQTENGE 1235

Qy   251 LMELLRFKEERILALEADMTKWEQ-------KYLEE----NVMRHFALDAA-------- 291
         || |:||||||||||:| ||: |:       ||||     |||||||||||
Db  1236 LARQLEEKEALTWQLTRGKLSYTQOMEDLKRQLEEEGKAKANALAH-ALQSARHDCDLLRE 1294

Qy   291 ------ATVAAQRDTIVISHSFN------TSYDT----------ALEARIQKEEE 323
               |: |:|||||||||||:      ||||           |||||||||||
Db  1295 QYEEEMEAKAELQR---VLSKANSFVAQWRTKYETDAIQRTEELEEAKKKLAQRLQDAEE 1351

Qy   324 EILMANKRCLDME-------GRIKTLHAQIIEKDAMIKVLQQORSRK----------EP 364
         :|| |:| |:|         |||| |||: |||||:|||| ||||:            |
Db  1352 AVEAVNAKCCSSLEKTKHRLQNEIEDLMVDVERSNAAAALDKKQRNFDKILAEWKQKYEE 1411

Qy   365 SKTEQLSCMRPAKS----LMSISNAGSGLLSHSSTLTGSPIMEEKRDDKS 410
         || |:| |:|| |     |: | | :| |:| :| |||:||||||| |:
Db  1412 SQSELESSQKEARSLSTELFKLKNAYEESLEHLETF------KRENKN 1453
```

RESULT 15

```
ID  W21731  standard; Protein; 2272 AA.
AC  W21731;
    31-OCT-1997  (first entry)
DE  GAL4/HA/NuMA fusion protein.
KW  NIP-1; NIP-2; NuMA; nuclear mitotic apparatus; NuMA interacting protein;
KW  cell division; proliferation; antibody; Ab; detection;
KW  malignant cell growth.
OS  Homo sapiens.
FH  Key            Location/Qualifiers
FT  domain         1..147
FT                 /label= GAL4_DNA_binding_domain
```

```
FT   peptide          148. .174
FT                    /label= Hemaglutinin_epitope
FT   protein          175. .2272
FT   07-JUN-1996;     U09504.
PF   19-DEC-1996;
PD   07-JUN-1995; US-478408.
PR   (UYYA ) UNIV YALE;
PA   McPherson SMG, Snyder MP;
PI   WPI; 97-077270/07.
DR   N-PSDB; T77782.
DR   New nucleic acid encoding nuclear mitotic appts. interacting
PT   proteins - useful for modulating cell division and proliferation and
PT   in diagnosis
PS   Claim 14; Page 28-36; 78pp; English.
CC   The sequences given in W21731-32 represent fusion proteins which contain
CC   NuMA (nuclear mitotic apparatus). The fusion proteins were used in
CC   the identification of NuMA interacting proteins (NIP's) (see also
CC   W21729-30). Compounds which interfere with the interaction of NuMA
CC   with a known NIP are used to modulate cell division and/or proliferation.
CC   Ab, raised conventionally using NIP-1 or -2 as immunogen, are used to
CC   detect NIP (or their complexes) and to block their activity for
CC   diagnostic or therapeutic use, e.g. to detect defective NuMA or NIP
CC   which may be markers for aberrant (including malignant) cell growth
CC   (which can also be detected by nucleic acid sequencing). Also where
CC   malignancy is related to defects in NuMA or NIP, it can be treated by
CC   administration of the appropriate functional protein.
SQ   Sequence 2272 AA;
```

FT   region   365. .1864
FT   /label= Residues_18-2116_of_NuMA
FT   region   /label= Coiled_coil_region

```
Query Match      6.1%;   Score 204;   DB 1;   Length 2272;
Best Local Similarity   21.2%;   Pred. No. 0.00022;
Matches  101;   Conservative  82;   Mismatches 163;   Indels 130;   Gaps  17;

Qy    24 RAQQMVEILSDENRNLRQELEGCYEXVARLQKVETEIQRVSEAYENLVKSSSKREALEKA  83
         ||||               |||          |:|            :||||   |
Db  1156 RGQQEREVARLTQERGRAQADLALEKAAR---AELEMRLQNALNEQRVEFATLQERALAHA 1212

Qy    84 MRNKLEG------EIRRMHDFN---------RDLRERLETANKQLAEKEYE----GS------ 122
         : ||       ||            ||    |:|:|||:|  |||:||   ||
Db  1213 LTEK-EGKDQELAKLRGLEAAQIKELELELRQTVKQLKEQLAKKEKEHASGSGAQSEAAGR 1271

Qy   122 -------EDTRKTISQLFAKNKESQREKEKLEAELATARSTN--------------EDQ  159
                |:||       |||||:|| || || ||          |:            :
Db  1272 TEPIGPKLEALRAEVSKLEQQCQKQQEQADSLERSLEAERASRAERDSALETLQGQLEEK 1331

Qy   160 RRHIEIRDQALSNAQ---------------AKVVKLEELKKQVYDRVEK--         197
         ||:||||          |:|  :||||:||   |:|
Db  1332 AQELGHSQSALASAQRELAAFRTKYQDHSKAEDEWKAQVARGQEAERKNSLISSLEEEV 1391

Qy   197 ----------MQQALVQ-----------------LQAAGEKREQLEHRLR--------------TRLEREL  226
                :|:|             |:|:                   ||||   ||
Db  1392 SILNRQVLEKEGESKELKRLVMAESEKSQKLEERLRLLQAETASNSARAERSSALREEV 1451

Qy   227 ESLRIQQRQGNCQFPNVSEYNAAALMELLREKEERIIALEADMTKWEQKYLE-ENVMRHF  285
         :|||||   :|||       |:||           :||    :|||||||:|  :|||||
Db  1452 QSLREEAEKQRVASEN-----LRQELTSQAERAEELGQELKAWQEKFFQKEQALSTL 1503

Qy   286 ALDAAATVAAQRDTTVISH--SPNTSYDTALEARIQKEEE----------EILMANKRC  332
         :|:|             ||    | :|::   |:|||||          |:||||||
Db  1504 QLEHTSTQALVSELLPAKIHCQQLQAEQAAAEKRHREELEQSKQAAGGLRAELLRAQREL 1563

Qy   333 LDMEGRIKTLHAQIIEKDAMIKVLQQRSRKEPSKTEQLSCMRPAKSIMSISNAGSG  388
         |:|| ||||:||   || |:|| ||     |||   || |||:| || || |
Db  1564 ----GELIPLRQKVAEQE---RTAQLRAEKASYAEQLSMLKKAHGLLAEENRGLG 1612
```

OM protein - protein search, using sw model

GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

Run On: January 18, 2000, 06:38:23 ; Search time 132.79 Seconds
                                               (without alignments)
                                               65.312 Million cell updates/sec

Title: US-09-332-063-2
Perfect score: 3347
Sequence: 1 MPRAQPSSASYQPVPADPFA..........KTPIQILGQEPDAEMVEYLI 675

Scoring table: BLOSUM62

Searched: 130275 seqs, 12848600 residues

Da se :   Issued_Patents_AA:*

Word size :   0

Number of hits that pass the threshold :   130275
    1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep:*
    2: /cgn2_6/ptodata/1/iaa/5B_COMB.pep:*
    3: /cgn2_6/ptodata/1/iaa/PCTUS9_COMB.pep:*
    4: /cgn2_6/ptodata/1/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	226.5	6.8	2482	1	US-08-328-254-6	Sequence 6, Appli
2	225	6.7	3248	1	US-08-353-700-1	Sequence 2, Appli
3	225	6.7	3248	3	PCT-US95-16216-1	Sequence 1, Appli
4	213	6.4	262	2	US-08-403-379A-1	Sequence 1, Appli
5	213	6.4	263	2	US-08-557-309B-51	Sequence 51, Appl
6	213	6.4	262	2	US-08-929-414-1	Sequence 1, Appli
7	209	6.2	219	2	US-08-557-309B-54	Sequence 54, Appl
8	208.5	6.2	1248	2	US-09-080-897-2	Sequence 2, Appli
9	204	6.1	2101	3	US-0-195-487-4	Sequence 4, Appli
10	204	6.1	2101	4	PCT-US93-06160-4	Sequence 4, Appli
11	203	6.1	885	2	US-08-533-306A-2	Sequence 2, Appli
12	203	6.1	816	2	US-08-533-306A-4	Sequence 4, Appli
13	203	6.1	816	2	US-08-533-306A-6	Sequence 6, Appli
14	203	6.1	576	2	US-08-742-923A-2	Sequence 2, Appli
15	203	6.1	885	2	US-08-742-923A-4	Sequence 4, Appli
16	203	6.1	816	2	US-08-742-923A-6	Sequence 6, Appli
17	203	6.1	180	4	US-273901-7	Sequence 6, Appli
18	203	6.1	180	4	Patent No. 5273901	Patent No. 5273901
19	201.5	6.0	2101	1	US-08-466-390-4	Sequence 4, Appli
20	201.5	6.0	2101	1	US-08-470-950-4	Sequence 4, Appli
21	201.5	6.0	2101	1	US-08-467-781-4	Sequence 4, Appli
22	201.5	6.0	2101	2	US-08-483-924-4	Sequence 4, Appli
23	199.5	6.0	1388	2	US-08-685-576-4	Sequence 4, Appli
24	199.5	5.9	955	2	US-08-428-414A-3	Sequence 3, Appli
25	199	5.9	1255	2	US-09-080-897-4	Sequence 4, Appli
26	197.5	5.9	1388	2	US-08-685-576-1	Sequence 1, Appli
27	193.5	5.8	955	1	US-08-006-676B-1	Sequence 1, Appli
28	193.5	5.8	955	2	US-08-282-845-1	Sequence 1, Appli
29	193.5	5.8	955	3	PCT-US94-00324-1	Sequence 1, Appli
30	191	5.7	1093	3	PCT-US94-03077-1	Sequence 1, Appli
31	187.5	5.6	486	1	US-09-450-360-2	Sequence 2, Appli
32	184.5	5.5	1898	1	US-08-056-200-94	Sequence 94, Appl
33	184.5	5.5	1898	2	US-08-800-644-94	Sequence 94, Appl
34	183	5.5	6.7	1	US-08-188-582-13	Sequence 13, Appl
35	183	5.5	677	1	US-08-646-715-13	Sequence 13, Appl

ALIGNMENTS

RESULT   1
US-08-328-254-6
Sequence 6, Application US/08328254
Patent No. 5710022
GENERAL INFORMATION:
APPLICANT: Zhu, Xueliang
APPLICANT: Lee, Wen-Hwa
TITLE OF INVENTION: A No. 5710022el Nuclear Mitotic Phosphoprotein
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
   ADDRESSEE: Campbell and Flores
   STREET: 4370 La Jolla Village Drive, Suite 700
   CITY: San Diego
   STATE: California
   ZIP: 92122
   COUNTF:Y: USA
COMPUTER READABLE FORM:
   MEDIUM TYPE: Floppy disk
   COMPUTER: IBM PC compatible
   OPERATING SYSTEM: PC-DOS/MS-DOS
   SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
   APPLICATION NUMBER: US/08/328,254
   FILING DATE: 24-OCT-1994
   CLASSIFICATION: 435
PRIOR APPLICATION DATA:
   APPLICATION NUMBER: US 08/141,239
   FILING DATE: 22-OCT-1993
ATTORNEY/AGENT INFORMATION:
   NAME: Campbell, Cathryn A.
   REGISTRATION NUMBER: 31,815
   REFERENCE/DOCKET NUMBER: P-CJ 1191
TELECOMMUNICATION INFORMATION:
   TELEPHONE: (619) 535-9001
   TELEFAX: (619) 535-8949
INFORMATION FOR SEQ ID NO: 6:
   SEQUENCE CHARACTERISTICS:
   LENGTH: 2482 amino acids
   TYPE: amino acid
   TOPOLOGY: linear
   MOLECULE TYPE: protein
US-08-328-254-6

Query Match   6.8%;   Score 226.5;   DB 1;   Length 2482;
Best Local Similarity   21.4%;   Pred No. 6.1e-07;
Matches 139; Conservative 100; Mismatches 233;   Indels 179;   Gaps 26;

Qy   26 QQMVEILSDSDNRNLRQE.E----------GCYEKVARLQKVETEIQRVSEAYENLVKSSSKRE 78
             : ::| :|| ||| ||||:|| : |: : :: |
Db   1948 QDTLEVIQSSYKNLENELELTKMDKMSFVEKVRKMTAKETELQR--EMHEMAQKTAELQE 2005

Qy   79 ALEKAMPNKLEGEIRRMHD--------------------FNRDLRERLETANKQLAEKEYEYGSED- 124
             : ||: | |: |:|:| : |: ||: |
Db   2006 EI-SGEKNRLAGELQLLLEEIKSSKDQLKELTLENSELKKSLDCMHKDQVEKEGKVREEI 2064

Qy   124 ------   --TRKIISQLFAKNESQREREKLEAELATARSTNEDQRRHIEI---RDQAL 170
                                   |: | :| : |: |

RESULT 2
US-08-353-700-1
Sequence 1, Application US/08353700
Patent No. 5599919
GENERAL INFORMATION:
APPLICANT: YEN, TIMOTHY J.
APPLICANT: RATTNER, JEROME B.
TITLE OF INVENTION: NUCLEIC ACID ENCODING A
TITLE OF INVENTION: TRANSIENTLY EXPRESSED KINETOCHORE PROTEIN,
TITLE OF INVENTION: AND METHODS OF USE
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: DANN, DORFMAN, HERRELL AND SKILLMAN
STREET: 1601 MARKET STREET, SUITE 720
CITY: PHILADELPHIA
STATE: PA
COUNTRY: USA
ZIP: 19103-2307
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/353,700
FILING DATE: 09-DEC-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: REED, JANET E.
REGISTRATION NUMBER: 36,252
TELECOMMUNICATION INFORMATION:
TELEPHONE: (215) 563-4100
TELEFAX: (215) 563-4044
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 3248 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: HUMAN
US-08-353-700-1

```
Db  2065  AEYQLRHEAEKKHQALLLDTNKQYEVELQTYREKLTSKEECLSSQKIEIDLLKSSKEEL  2124
Qy   171  SNAQAKVVKLEEELKKQV-----YVDKV-----EKMQQALVQLQQAACEKREQLEHRLRTRL  222

Db  2125  NNSLKATTQILEELKTKMDNLKYVNQLKKENERAQGKMKLLIKSCKQLEEKEILQKEL  2184
Qy   223  ERELESLRIQQRQGNCQQPTNVSEYNA--AAIMELLREKEERILALEADMTKWEQKYLEEN  280

Db  2185  S-QLQAAQEKQKTGTVMDTRVDELITEIKELKETLEEK---------TKEADEYLDKY  2232
Qy   281  VMRHFALDAAATVAAQRDTTVISHSPNTSYDTALEARIQKEEEEILMANKRCLDMEGRIK  340

Db  2233  C--------SLLISHE--------KLEKAKEM------LETQVA  2254
Qy   341  TLHAQIIEKDA-----MIKVLQQRSREPSKTEQLSCMRPAKSLMSISNAGSGLLSHSSTL  396

Db  2255  HLCSQQSKQDSRGSPLLGPVVPGPSPIPSVTE------KRLSSGQNRASGKRQRSSGI  2306
Qy   397  TGSPIMEEKRDDKSWKGSLGILLGGDYRAEYVPSTPSPVPPSTPLLSAHSKTGSRDCSTQ  456

Db   307  ----------WENG----GG------PTPATPESFSKKSKSKKAVMSGIHPAE  2337

Qy   457  TERGTESNKTAAVAPISVPAPVAAAATAAAITATPAATI--TTTMVAAPVAVAAAAAPAA  514
Db  2338  DTEGTE-----FEPEGLPEVVKKGFADIPTGKTSPYILRRTTM------  2376

Qy   515  AAAPSPATANAATANAAVSPAAAAGQIPAAASVASAAAAVAPSAAAAAAVQVAPAAAPAPVPAPA  574
Db  2376  ATRTSP-RLAAQKLALSPLSLGKENLAES-----SKPTAGGSRSQKVKVAQRSPVDSGT  2428

Qy   575  LVPVPAPAAAQASAPAQTQAPTSAPAVAPTPAPTPTPAVAQAEVPASPATG  625
Db  2429  ILREP----TTKSVPV-----NNLPERSPTDSPREGLRVKRGRLVPSPKAG  2470
```

RESULT 3
PCT-US95-16216-1
Sequence 1, Application PCT/US9516216
GENERAL INFORMATION:
APPLICANT: Yen, Timothy J.
APPLICANT: Rattner, Jerome B.
TITLE OF INVENTION: Nucleic Acid Encoding a Transiently
TITLE OF INVENTION: Expressed Kinetochore Protein, and Methods of Use
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: Dann, Dorfman, Herrell and Skillman
STREET: 1601 Market Street, Suite 720
CITY: Philadelphia
STATE: PA
COUNTRY: USA
ZIP: 19103-2307
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/16216

Query Match 6.7%; Score 225; DB 1; Length 3248;
Best Local Similarity 21.4%; Pred. No. 1.1e-06;
Matches 110; Conservative 106; Mismatches 167; Indels 132; Gaps 21;

```
Qy    20  AIVSRAQQMYEILLSDENRNLRQELEGGCYEKVARLQK--------VETEIQRVSEAYE   68
Db  2103  AEVKEKTELLQTLSSDVSELLKDKTHLQEKLQSLEKDSQALSLTKCELENQIAQLNKEKE  2162

Qy    69  NLVKSSSK--------REALEKAMRNKLEGEIR-----RMHDFNR---DLRE   104
Db  2163  LLVKESESLQARLSESDYEKLNVSKALEAALVEKGEFALRLSSSTQEEVHQLRGIEKLRV  2222

Qy   105  RLETANKQ--LAEK-------EYEGSEDTRKTISOLFAKNRESQR----EKEKLEAELAT  151
Db  2223  RIEADEKKQLHIAEKLKEREREND SLKDKVENLERBLQMSEENQELVILDAENSKAEVET  2282

Qy   152  ARSTNEDQRRHIEIRDQALLSNAQAKVVKLEEELKKKQVYVDKVEKM-------QQA  200
Db  2283  LKTQIEEMARSLKIFELDLVTLRSEKENLTKQIQEKQGQLSELDKLLSSFKSLLEEKEQA  2342

Qy   201  LVQLQ-------AACEKRE-----QLEHRLRTRLERE  225
Db  2343  EIQIKEESKTAVEMLQNQLKELNEAVAALCGQQEIMKATEQSLDPPIEEEHQLRNSIEKL  2402

Qy   226  LESLRIQQRQGNC--QPTNVSEYNNAAL-------MELLREKEERIIALEADMTKWE  273
Db  2403  RARLEADEKKQLCVLQQLRESEHRADLLKGRVENLERELEIARTNQEH-AALEAENSKGE  2461

Qy   274  QKYLEENV-----MRHFALDAAA-----TVAAQRDTTVISHSP--NTSYDTALEAR  317
Db  2462  VETLKAKIEGMTQSLRGLELDVVTIRSEKENLTNELQKEQRISELEINSSFENILQ--  2520

Qy   318  IQKEEEEILMANKRCLDMEGRIKTLHAQIIEKDAMIKVLQQRSR-REPSKTEQLSC--MR  374
Db  2520  -EKEQEKYQMKEKSSTAME-MLQTQLKELNERVAALHNDQEACAKEQNLSSQVECLELE  2577

Qy   375  PAKSLMSISNAGSGLLSHSSTLTGSPIMEEKRDDK  409
Db  2578  KAQLLQGLDEAKNNYIVLQSSVKG--LIQQEVEDGK  2610
```

FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/353,700
FILING DATE: 09-DEC-1995
ATTORNEY/AGENT INFORMATION:
NAME: Reed, Janet E.
REGISTRATION NUMBER: 36,252
TELECOMMUNICATION INFORMATION:
TELEPHONE: (215) 563-4100
TELEFAX: (215) 563-4044
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 3248 amino acids
TYPE: amino acid
STRANDEDNESS: not relevant
TOPOLOGY: not relevant
MOLECULE TYPE: protein
HYPOTHETICAL: NO
ANTI-SENSE: NO
PCT-US95-16216-1

Query Match   6.7%; Score 225; DB 3; Length 3248;
Best Local Similarity 21.4%; Pred. No. 1.1e-06;
Matches 110; Conservative 106; Mismatches 167; Indels 132; Gaps 21;

Qy 20 AIVSRAQQMVEILSDENRNLRQELEBGCYEKVARLQK-----------VETEIQRVSEAYE 68
        :|||| |: | |: |:   ::| | ||: | ||
Db 2103 AEVKEKTELLQTLSSDVSELLKDKTHLQEKLQSLEKDSQALSLTKCELENQIAQLNKEKE 2162

Qy 69 NLVKSSSK--------------REPLIEKAMRNKLEGEIR-----RMIDFNR---DLRE 104
        ||| |||| |: || :| : || | : ||
Db 2163 LIVKESESLQARLSESDYEKLNVSKALEAALVEKGEFALRLSSTQEEVHQLRGIEKLRV 2222

Qy 105 RLETANKQ---LAEK------EYEGSEDFRKTISQLFAKNKESQR----EKEKLEAELAT 151
        |: | ::| ||:| : || :|:| | :|:| |
Db 2223 RIEADEKKQLHIAEKLKERERENDSLKDKVENLERELQMSEENQELVILDAENSKAEVET 2282

Qy 152 ARSTNEDQRRHEIRDQALSNAQAKVKVKLEEELKKQVYDKVEKM----------QQA 200
        :|| :| | | :| || : |:|: |: | ||
Db 2283 LKTQIEEMARSLKIFELDLVTLRSEKENLTKQIQEKQGQLSELDKLLSSFKSLLEEKEQA 2342

Qy 201 LVQLQ-------------AACEKRE----------QLEHRLRTRLERE- 225
        :||:   :| : |: |:|:|:|:
Db 2343 EIQIKEESKTAVEMLQNOLKELNEAVAALCGDQEIMKATEQSLDPPIEEEHQLRNSIEKL 2402

Qy 226 LESLRIQORQGNC--QPTNVSEYINAAAL-------------MELLREKEERILALEADMTKWE 273
        ::: ||| ||| | |||:|| :||:|
Db 2403 RARLEADEKKQLCVLQOLKESEHHADLLKGRVENLERELEIARTNQEH-AALEAENSKGE 2461

Qy 274 QKYLEENV--------MRHFALDAAA----------TVAAQRDTTVISHSP--NTSYDTALEAR 317
        |: | |: | ||:|::| :|:: |
Db 2462 VETLKAKIEGMTQSLRGLELDVVTIRSEKENLTNELQKEQERISELEINSSFENILQ-- 2520

Qy 318 IQKEEEIIMANKRCLDMEGRIKTLHAQIIEKDAMIKVLQQRSR-KEPSKTEQLSC--MR 374
        :||:|:| ||| | |:   ::| |: |||: : || : |
Db 2520 -EKQEKVQMKEKSSTAME-MLQTOLKELNERVAALHNDQEACKAKEQNLSSQVECLELE 2577

Qy 375 PAKSLMSISNAGSGLLSHSSTLTGSPIMEEKRDDK 409
        :   :: |:   :| :: || |
Db 2578 KAQLLQGLDEAKNNYIVLQSSVKG--LIQEVEDGK 2610

RESULT   4
US-08-403-379A-1
Sequence 1, Application US/0840337 9A
Patent No. 5756662
GENERAL INFORMATION:
APPLICANT: Reed, Steven G.
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DETECTION
TITLE OF INVENTION: OF T. CRUZI INFECTION
NUMBER OF SEQUENCES: 5

CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED and BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/403,379A
FILING DATE: 14-MAR-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Maki, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.406
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 262 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-403-379A-1

Query Match   6.4%; Score 213; DB 1; Length 262;
Best Local Similarity 48.4%; Pred. No. 2.8e-07;
Matches 74; Conservative 9; Mismatches 54; Indels 16; Gaps 9;

Qy 479 AAAAATAAAIITATAAGITTTTMVAAAP---VAVAAAAAPA-AAAAPSPATAAAATAAAAVSPAA 534
         ||| |||||| |||||||||:| || :||: ||||| ||||:|
Db 112 AAAAAAKQKAAAAK-----AAAPSGKKSAKAAIAPAKAAAAPAKAAAAPAKAAAAAPAK 165

Qy 535 AGQIPAAASVASA-AAVAPSAAAAAAVQVAPAAAPAVPAPRLVPVPAPAAAQASAPAQTQ 593
         || ||: |:| |||:|| |||| |||||||| |||||||||| |||: | |||
Db 166 AAAAPAKAAAAPAKAAATAPAKAAAAPAKAAAAPAKTA-AAPAKAAAAPAKA-AAAPAKA-ATAPAKAA 222

Qy 594 AFTSAPAVAPTPAPTPTPAVAQAEVPASPATGP 626
         ||| |||||| :||| | |: |||||||
Db 223 AAPAKAATAPAKAAT-APAKA-AAAPAKAATAP 253

RESULT   5
US-08-557-309B-51
Sequence 51, Application US/08557309B
Patent No. 5916572
GENERAL INFORMATION:
APPLICANT: Reed, Steven G.
APPLICANT: Skeiky, Yasir A.W.
APPLICANT: Lodes, Michael J.
APPLICANT: Houghton, Raymond L.
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DETECTION AND PREVENTION OF
NUMBER OF SEQUENCES: 69
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED and BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:

(Left panel)

```
APPLICATION NUMBER: US/08/557,309B
FILING DATE: 14-NOV-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Maki, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.422
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 51:
SEQUENCE CHARACTERISTICS:
LENGTH: 263 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear

US-08-557-309B-51
```

`^ Y Match`
```
Local Similarity     6.4%;     Score 213;   DB 2;
M..ches    74;  Conservative   9;  Pred. No. 2.9e-07;
                            Mismatches   54;   Indels   16;   Gaps   9;
```

```
Qy    479  AAAATAAAITATATAITTTTMVAAAP---VAVAAAAAPA-AAAAPSPATAAATAAAVSPAA 534
           ||||| ||    | |: :||   |||| |||| |||||:  | |||||| ||:|| |||:||
Db    113  AAAAAAAKQKAAAKK------AAAPSGKKSAKAAIAPAKAAAAAPAKAAAAAPAKAAAAPAK 166

Qy    535  AGQIPAAASVASA-AAVAPSAAAAAAVQVAPAPAPVPAPALVPVPAPAAAQASAPAQTQ 593
           | ||:| |||    ||||: || : |||| |||| ||||   ||| |:|:|| |||:|
Db    167  AAAAPAKAAAAPAKAAATAPAKAAAAAPAKTA-AAPAKAAAAPAKA-AAAPAKA-ATAPAKAA 223

Qy    594  APTSAPAVAPTPAPTPTPAVAQAEVPASPATGP 626
           |:| ||:| | || |||| || | ||| ||| |
Db    224  AAPAKAATAPAKAAT-APAKA-AAAPAKAATAP 254
```

```
RESULT   6
US-08-929-414-1
Sequence 1, Application US/08929414
Patent No. 594203
GENERAL INFORMATION:
APPLICANT: Reed, Steven G.
APPLICANT: Haughton, Raymond
APPLICANT: Skeiky, Yasir A.W.
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DETECTION
TITLE OF INVENTION: OF T. CRUZI INFECTION
NUMBER OF SEQUENCES: 15
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED and BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/929,414
FILING DATE: 15-SEP-1997
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: Maki, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.406C1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 262 amino acids
```

(Right panel)

```
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein

US-08-929-414-1
```

`Query Match`
```
Best Local Similarity   6.4%;   Score 213;   DB 2;
Matches   74;   Conservative   9;   Pred. No. 2.8e-07;
                            Mismatches   54;   Indels   16;   Gaps   9;
```

```
Qy    479  AAAATAAAITATATAITTTTMVAAAP---VAVAAAAAPA-AAAAPSPATAAATAAAVSPAA 534
           ||||| ||    | |: :||   |||| |||| |||||:  | |||||| ||:|| |||:||
Db    112  AAAAAAAKQKAAAKK------AAAPSGKKSAKAAIAPAKAAAAAPAKAAAAAPAKAAAAPAK 165

Qy    535  AGQIPAAASVASA-AAVAPSAAAAAAVQVAPAPAPVPAPALVPVPAPAAAQASAPAQTQ 593
           | ||:| |||    ||||: || : |||| |||| ||||   ||| |:|:|| |||:|
Db    166  AAAAPAKAAAAPAKAAATAPAKAAAAAPAKTA-AAPAKAAAAPAKA-AAAPAKA-ATAPAKAA 222

Qy    594  APTSAPAVAPTPAPTPTPAVAQAEVPASPATGP 626
           |:| ||:| | || |||| || | ||| ||| |
Db    223  AAPAKAATAPAKAAT-APAKA-AAAPAKAATAP 253
```

```
RESULT   7
US-08-557-309B-54
Sequence 54, Application US/08557309B
Patent No. 5916572
GENERAL INFORMATION:
APPLICANT: Reed, Steven G.
APPLICANT: Skeiky, Yasir A.W.
APPLICANT: Lodes, Michael J.
APPLICANT: Houghton, Raymond L.
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DETECTION AND PREVENTION OF
NUMBER OF SEQUENCES: 69
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED and BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/557,309B
FILING DATE: 14-NOV-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Maki, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.422
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 54:
SEQUENCE CHARACTERISTICS:
LENGTH: 219 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear

US-08-557-309B-54
```

`Query Match`
```
Best Local Similarity   6.2%;   Score 209;   DB 2;
Matches   69;   Conservative   6;   Pred. No. 4.2e-07;
                            Mismatches   71;   Indels   12;   Gaps   5;
```

```
Qy    476  APVAAAAAAAAITATAAITTTTMVAAAPVAVAAAAAPAAAAAAAPSPATAAATA------A 528
           || |   |   || |  :||   |||||| ||||||: |||||| |||| ||||     |
Db     58  AAAAAKQKAAAKAAAPSGKKSAKAAAP--AKAAAAPAKAAAAPAKTAAAPAKAAAAHAKA 115
```

Qy   529 AVSPAAAGQIPAAASVASAAAVAPSAAAAAVQVAPARAPAPVPAPALVPPARPAAAQASA 588
           | ||| | |||| || || ||||||| ||| | |||||| ||||
Db   116 AAPPAKAAAPPAKTAAPPAKTRAPPAKAAAAPPAKAAAAPPAKAAAAPPAKAAAAPAKA-AAA 174

Qy   589 PAQTQAPTSAPAVAPTPAPTPTPAVAQAEVPASPATGP 626
           | ||| ||||| |||| || | | |||||| || |||| |
Db   175 PAKAAAPPAKAAAPPAKAAAP-PAKAAAP-PAEAAPAP 210

RESULT 8
US-09-080-897-2
Sequence 2, Application US/09080897
Patent No. 5985574
GENERAL INFORMATION:
    APPLICANT: King, Mary-Claire
    APPLICANT: Lynch, Eric D.
    APPLICANT: Lee, Ming
    PPLICANT: Morrow, Jan E.
    PPLICANT: Welsh, Piri L.
    APPLICANT: Leon, Pedro E.
    TITLE OF INVENTION: Modulators of Actin
    NUMBER OF SEQUENCES: 14
    CORRESPONDENCE ADDRESS:
        ADDRESSEE: SCIENCE & TECHNOLOGY LAW GROUP
        STREET: 75 DENISE DRIVE
        CITY: HILLSBOROUGH
        STATE: CALIFORNIA
        COUNTRY: USA
        ZIP: 94010
    COMPUTER READABLE FORM:
        MEDIUM TYPE: Floppy disk
        COMPUTER: IBM PC compatible
        OPERATING SYSTEM: PC-DOS/MS-DOS
        SOFTWARE: Patentin Release #1.0, Version #1.30
    CURRENT APPLICATION DATA:
        APPLICATION NUMBER: US/09/080,897
        FILING DATE:
        CLASSIFICATION:
    ATTORNEY/AGENT INFORMATION:
        NAME: OSMAN, RICHARD A.
        REGISTRATION NUMBER: 36,627
        REFERENCE/DOCKET NUMBER: UW97-001
    TELECOMMUNICATION INFORMATION:
        TELEPHONE: (650) 343-4341
        TELEFAX: (650) 343-4342
    SEQUENCE INFORMATION FOR SEQ ID NO: 2:
        SEQUENCE CHARACTERISTICS:
        LENGTH: 1248 amino acids
        TYPE: amino acid
        TOPOLOGY: linear
    MOLECULE TYPE: protein
US-09-080-897-2

Query Match   6.2%;   Score 208.5; DB 2;   Length 1248;
Best Local Similarity   20.4%;   Pred. No. 4e-06;
Matches 162; Conservative 95; Mismatches 274; Indels 263; Gaps 39;

Qy   7 SSASYQPVP-ADPFAIVSRAQOMV------EILSDENRN----------LRQELEGGCYEK 49
           | ||| | | | | | ||| | |: | | | | | ||| |
Db   62 SSASYGDDPTPAQSLQDVDSDCQVLVLFEQMLLDMNLNEEKQQPLREKDIIIKREMVSQYLY 121

Qy   50 VARLQKVETEIQRVSEAYENLVRSS-------SKREALEKAMRN------------K 87
           | :   : | | | | | | : |
Db   122 TSKAGMSQKESSKSAMYIQELRSGLRDMPLLSCLESLRVSLNNNPVSWVQTFGAEGLAS 181

Qy   88 LEGEIRRMHDFNRDLRERLETANKQLAEKEYEGSEDTRK----TISQLFAKNKESQREK 142
             ::| :| | ||| | ||| |
Db   182 LLDLIKRLHD------FKEETA-------GSYDSRNKHEIIRCLKAFMNKFG---- 222

Qy   143 EKLEAELATARSTNEDQRRHIEIRDQALSNAQAKVVKL---------EEELKKK------ 188
           | | |: | :: :
Db   182 ... 188

Db   222 ------IKTMLETTGILILVRAMDPAVPNMMIDAAKLLSALCILPQPEDMNERVLEAMT 275
                       |: ||| | |
Qy   188 -QYVDKVYKMQQAL-------VQLQAAC-------EKREQLEHRLRTRLE------- 224

Db   276 ERLEMDEVERFQPLLDGLKSGTTIALKVGCLQLINALITPAEELDFRVHIRSELMRLGLH 335
              | : | :| | | |: | : | | | |: |: |: | | |
Qy   224 ------RELES---LRIQ---QRQGNCQP--------TNVSEYNAA--ALMELLREK 258

Db   336 QVLQDIRELENEDMRVQLNVFDEQGEEDSYDLKGRLDDIRMEMDDFNEVFQILLNTVKDS 395
             | | |: | |: ::
Qy   259 E---------ERILALEADMTKWEQKY--LEENV----------MRHFALDAAT 292

Db   396 KAEEHFLSILQHILLVRNDYEARPQYYKLIEECISQIVLHKNGADPDFKCRHLQIEIEGL 455
           | || | | | | | | | |: ||| | | | | | | |
Qy   293 VAAQRDTTVISHSP------NTSYDTALEARIQKEEEILMANKRCLDMEGRIKTLHAQI 346

Db   456 IDQMIDKTKVEKSEAKAAELEKKLDSELTARHELQVEMKKMES----DFEQKLQDLQG-- 510
           | | : ||| | || || | | | | | :   : | | | |
Qy   347 IEXLAMIKVLQQRSRKEPSKTFQLSCMRPAKSLMSISNAGSGLLSHSSTLTG--SPIMEE 404

Db   510 -EKDAL------HSEKQOIATEK-------------QDLEAEVSQLTGEVAKLTKE 545
             :| | |: |: | |: | |: | |: | |
Qy   405 KPPDKSWKGSLGTLLGGDYRAEVVPSTPS--PVPPSTPLLSAHSKTGSRDCSTQTERGTE 462

Db   546 LEDAKKEMASLSAA------AIVPPSVPSRAPVPPAPL---------PGDSGT- 586
           | | | | : | : | | ||: | || |
Qy   463 SNKTAAVAPISVPAPVAAAATAAAITATAATITTTMVAAAPVAVAVAAAAP--AAAAAPSP 520

Db   586 --------IIPPPAPGD---------STTPPPPPPPPPPPPPLPGGGTAISPPP 623
                 | || | | |||||||||||||| || || |||
Qy   521 ATAAATAAAYSPAAG-QIPAAASVASAAAAVAPSAAAAAVQVAPARAPAPVPAPA-LVPV 578

Db   624 LSGDATIPPPPPLPEGVGIPSPSSLPGGTAIPPPPPLPGSARI-PPPPPLPGGSAGIPPP 682
           | | || ||||| || | | | | ||||||||| |||||||| ||||
Qy   579 PAPAAAQAASAPAQTQAPTSAPAVAPTP------APTPTPAVAQAE----VPASPAT 624

Db   683 PPPLPGEAGMPPPPPLPGGPGIPPPPPPFGGGPGIPPPPPGGPMGMPPPPPFGFGVPAAPVL 742
                 | | || || | | || | | ||: |
Qy   625 --GPGHRISIPSL 636

Db   743 PFGLTPKLIKPEV 756

RESULT 9
US-08-195-487-4
Sequence 4, Application US/08195487
Patent No. 5783403
GENERAL INFORMATION:
    APPLICANT: TOUKATLY, GARY
    APPLICANT: LIDGARD, GRAHAM F
    TITLE OF INVENTION: NOVEL MALIGNANT CELL TYPE MARKERS OF THE
    TITLE OF INVENTION: INTERICK NUCLEAR MATRIX
    NUMBER OF SEQUENCES: 6
    CORRESPONDENCE ADDRESS:
        ADDRESSEE: TESTA HURWITZ & THIBEAULT
        STREET: 53 STATE STREET
        CITY: BOSTON
        STATE: MA
        COUNTRY: USA
        ZIP: 02109
    COMPUTER READABLE FORM:
        MEDIUM TYPE: Floppy disk
        COMPUTER: IBM PC compatible
        OPERATING SYSTEM: PC-DOS/MS-DOS
        SOFTWARE: Patentin Release #1.0, Version #1.25
    CURRENT APPLICATION DATA:
        APPLICATION NUMBER: US/08/195.487
        FILING DATE:
        CLASSIFICATION:
    PRIOR APPLICATION DATA:
        APPLICATION NUMBER: 435
        APPLICATION NUMBER: US/07/901.701
        FILING DATE:
    ATTORNEY/AGENT INFORMATION:

RESULT 10
PCT-US93-06160-4
Sequence 4, Application PC/TUS9306160
GENERAL INFORMATION:
APPLICANT:
TITLE OF INVENTION: NOVEL MALIGNANT CELL TYPE MARKERS OF THE
TITLE OF INVENTION: INTERIOR NUCLEAR MATRIX
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: TESTA HURWITZ & THIBEAULT
STREET: 53 STATE STREET
CITY: BOSTON
STATE: MA
COUNTRY: USA
ZIP: 02109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:

NAME: PITCHER ESQ, EDMUND R
REGISTRATION NUMBER: 27,829
REFERENCE/DOCKET NUMBER: MTP-013
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617/248-7000
TELEFAX: 617/248-7100
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 2101 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-195-487-4

Query Match 6.1%; Score 204; DB 1; Length 2101;
Best Local Similarity 22.3%; Pred. No. 1.5e-05;
Matches 106; Conservative 78; Mismatches 163; Indels 128; Gaps 19;

```
Qy    24 RAQQMVEILSDENRNLRQELEGCYEKVARLQKVETEIQRVSEAYENLVKSSSKREALEKA 83
Db   999 RGQQEREVARLTQERGRAQADLALEKAAR---AELEMRLQNALNEQRVEFATLQEALAHA 1055

Qy    84 MRNKLEG------EIRRMHDFN------RDLRERLETANKQLAEKEYE---GS------ 122
Db  1056 LTEK-EGKDQELAKLRGLEAAQIKELEELRQTVKQLKEQLAKKEKEHASGSGAQSEAAGR 1114

Qy   122 ------EDTRKIISQLFAKNKESQREKEKLEAELATARSTNEDQRRHIEIRDQALSNA 173
Db  1115 TEPTGPKLEALRAEVSKLEQQCQKQOEQADSLERSLEAERASRAE------RDSALETL 1167

Qy   174 -----VYVDKVEKMQQALVQLQAACEK-REQLEHR--LR 219
Db  1168 QGQLEEKAQELGHSQSALASAQERELAAFRTKVQDHSKAEDEWKAQVARGRQEAERKNSLI 1227

Qy   220 TRLREL------ESLRIQQRQGNC-----OPTNVSEYNAAA 250
Db  1228 SSLEEEVSILNRQVLEKEGESKELKRLVWAESEKSQKLEESCACCRQROPATVPELQNAA 1287

Qy   251 LM----------ELLRE----KEERILALEADMTKWEQKYLE-ENVMRHFA 286
Db  1288 LLCGRRCRASGREAEKQRVASENLRQELTSQAERAEELGQELKAWQEKFFQKEQALSTLQ 1347

Qy   287 LDAAATVAAQRDTTVISH--SPNTSYDTFALEARIQKEEE------EILMANKRCL 333
Db  1348 LEHTSTQALVSELLPAKHLCQQLQAEQAAAEKHHREELEQSKQAAGGLRAELLRAQREL- 1407

Qy   334 DMEGRIKTLHAQIIEKDAMIKVLQORSRKEPSKTEQLSCMRPAKSLMSISNAGSG 388
Db  1407 --GELIPLRQKVAEQD---RTAQOLRAEKASYAEQLSMLKKAHGLLAEENRGLG 1455
```

Query Match 6.1%; Score 204; DB 3; Length 2101;
Best Local Similarity 22.3%; Pred. No. 1.5e-05;
Matches 106; Conservative 78; Mismatches 163; Indels 128; Gaps 19;

RESULT 11
US-09-533-306A-2
Sequence 2, Application US/08533306A
Patent No. 5837457
GENERAL INFORMATION:
APPLICANT: Liu, Pu
APPLICANT: Collins, Francis S.
APPLICANT: Sicilliano, Michael J.
APPLICANT: Claxton, David
TITLE OF INVENTION: Markers for Detection of Chromosome 16
TITLE OF INVENTION: Rearrangements
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:
ADDRESSEE: Harness, Dickey & Pierce, P.L.C.
STREET: P.O. Box 828
CITY: Bloomfield Hills
STATE: MI

```
Qy    24 RAQQMVEILSDENRNLRQELEGCYEKVARLQKVETEIQRVSEAYENLVKSSSKREALEKA 83
Db   999 RGQQEREVARLTQERGRAQADLALEKAAR---AELEMRLQNALNEQRVEFATLQEALAHA 1055

Qy    84 MRNKLEG------EIRRMHDFN------RDLRERLETANKQLAEKEYE---GS----- 122
Db  1056 LTEK-EGKDQELAKLRGLEAAQIKELEELRQTVKQLKEQLAKKEKEHASGSGAQSEAAGR 1114

Qy   122 ------EDTRKIISQLFAKNKESQREKEKLEAELATARSTNEDQRRHIEIRDQALSNA 173
Db  1115 TEPTGPKLEALRAEVSKLEQQCQKQOEQADSLERSLEAERASRAE------RDSALETL 1167

Qy   174 -----VYVDKVEKMQQALVQLQAACEK-REQLEHR--LR 219
Db  1168 QGQLEEKAQELGHSQSALASAQERELAAFRTKVQDHSKAEDEWKAQVARGRQEAERKNSLI 1227

Qy   220 TRLREL------ESLRIQQRQGNC-----OPTNVSEYNAAA 250
Db  1228 SSLEEEVSILNRQVLEKEGESKELKRLVWAESEKSQKLEESCACCRQROPATVPELQNAA 1287

Qy   251 LM----------ELLRE----KEERILALEADMTKWEQKYLE-ENVMRHFA 286
Db  1288 LLCGRRCRASGREAEKQRVASENLRQELTSQAERAEELGQELKAWQEKFFQKEQALSTLQ 1347

Qy   287 LDAAATVAAQRDTTVISH--SPNTSYDTFALEARIQKEEE------EILMANKRCL 333
Db  1348 LEHTSTQALVSELLPAKHLCQQLQAEQAAAEKHHREELEQSKQAAGGLRAELLRAQREL- 1407

Qy   334 DMEGRIKTLHAQIIEKDAMIKVLQORSRKEPSKTEQLSCMRPAKSLMSISNAGSG 388
Db  1407 --GELIPLRQKVAEQD---RTAQOLRAEKASYAEQLSMLKKAHGLLAEENRGLG 1455
```

Left column

```
;   COUNTRY:  USA
;      ZIP:  48303
;  COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;  CURRENT APPLICATION DATA:
;      APPLICATION NUMBER: US/08/533,306A
;      FILING DATE:  September 25, 1995
;      CLASSIFICATION:  435
;  ATTORNEY/AGENT INFORMATION:
;      NAME:  Smith, DeAnn F.
;      REGISTRATION NUMBER:  36683
;      REFERENCE/DOCKET NUMBER: 2115-00869COB
;  TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (810) 641-1600
;      TELEFAX:  (810) 641-0270
;  INFORMATION FOR SEQ ID NO: 2:
;  SEQUENCE CHARACTERISTICS:
;      LENGTH:  576 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;      MOLECULE TYPE:  protein

US-08-533-306A-2

Query Match             6.1%;  Score 203;  DB 2;  Length 576;
Best Local Similarity   23.4%;  Pred. No. 3.5e-06;
Matches  101;  Conservative  74;  Mismatches 150;  Indels 106;  Gaps  16;

Qy    30 EILSDENRNLR-----QELEGGYEKVARLQKVETE------IQRVSEAYENLVKSSSKREA 79
         |: :|| ||         || | :   |  |:  ||    :||       :  |     :| |
Db   192 ELQASEDAKLRLEVNMQALKGQFERDLQARDEQNEEKRRQLQROLHEYETELEDERNERA 251

Qy    80 LEKAMRNKLEGEIRRMHDFNRDLRERLETANKQLAE-----KEYEGS-EDTRKTISQLFA 133
         | |||| ::| |||||:::  | ||||||    ||       | |||  ||:|| : ||
Db   252 LAAAAKKLEGDLKDLELQADSAIKGREEAIKQLRKLQAQMKDFQRELEDARASRDEIFA 311

Qy   134 KNKESQR----------------EKEKLEAELATA---RSTNEDQRRH 162
         || ::|                  |||:|  || |   ::| |: ||
Db   312 TAKENEKKAKSLEADLMQLQEDLAAAERARKQADLEKEELAEELASSLSGRNALQDEKR 371

Qy   163 IEIRDQALSNAQAKVVKLEEELKKKQ-----VYVDKVEKMQQALVQL--QAACEK-REQLE 215
         |||| :  | |  |  :||||||||       | : | :|      |     |  ||:||
Db   372 LE---------ARIAQLEELEEQGNMEAMSDRVRKATQQAEQLSNELATERSTAQKN 421

Qy   216 HRLRTRLERELESLRIQORQGNCQPTNVSEYNAAALMELLREKEERILALEADMTKWEQK 275
         |: || | | :   |   ||    |  | :         |   | :| :| |  |:|:|
Db   422 ESARQOLERONKELR--------SKLHEMEGAVKSKF---KSTIAALEAKIAQ- 464

Qy   276 YLEENVMRHFALDAAATVAAQRDTTVISHSPNTSYDTALEARIQKEEEEILMANKRCLDM 335
         | :|    |: |  |         |:|:| :        |    | | |    ::| |
Db   464 -LEEQVEQEAREKQAATKSLKQK----------------DKKLKEILLQVEDERKMAEQYKEQA 510

Qy   336 E---GRIKTLHAQIIEKDAMIKVLQQRSRKEPSKTEQLSCMRPAKSLMSISNAGSGLLSH 392
         :   ||  ||        | |: | |
Db   511 EKGNARVKQLKRQLEEAEEESQRINANRKLQRELDEATESNEA------MGREVNAL 562

Qy   393 SSTLTGSPIME 4C3
         |: |:| |  |
Db   563 KSKLRGPPPQE 573
```

Right column

```
;   TITLE OF INVENTION:  Markers for Detection of Chromosome 16
;   TITLE OF INVENTION:  Rearrangements
;   NUMBER OF SEQUENCES:  14
;   CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Harness, Dickey & Pierce, P.L.C.
;      STREET:  P.O. Box 828
;      CITY:  Bloomfield Hills
;      STATE:  MI
;      COUNTRY:  USA
;      ZIP:  48303
;  COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;  CURRENT APPLICATION DATA:
;      APPLICATION NUMBER: US/08/533,306A
;      FILING DATE:  September 25, 1995
;      CLASSIFICATION:  435
;  ATTORNEY/AGENT INFORMATION:
;      NAME:  Smith, DeAnn F.
;      REGISTRATION NUMBER:  36683
;      REFERENCE/DOCKET NUMBER: 2115-00869COB
;  TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (810) 641-1600
;      TELEFAX:  (810) 641-0270
;  INFORMATION FOR SEQ ID NO: 4:
;  SEQUENCE CHARACTERISTICS:
;      LENGTH:  885 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;      MOLECULE TYPE:  protein

US-08-533-306A-4

Query Match             6.1%;  Score 203;  DB 2;  Length 885;
Best Local Similarity   23.4%;  Pred. No. 6e-06;
Matches  101;  Conservative  74;  Mismatches 150;  Indels 106;  Gaps  16;

Qy    30 EILSDENRNLR--:--QELEGGYEKVARLQKVETE------IQRVSEAYENLVKSSSKREA 79
         |: :|| ||        || | :   |  |:  ||    :||       :  |     :| |
Db   501 ELQASEDAKLRLEVNMQALKGQFERDLQARDEQNEEKRRQLQROLHEYETELEDERNERA 560

Qy    80 LEKAMRNKLEGEIRRMHDFNRDLRERLETANKQLAE-----KEYEGS-EDTRKTISQLFA 133
         | |||| ::| |||||:::  | ||||||    ||       | |||  ||:|| : ||
Db   561 LAAAAKKLEGDLKDLELQADSAIKGREEAIKQLRKLQAQMKDFQRELEDARASRDEIFA 620

Qy   134 KNKESQR----------------EKEKLEAELATA---RSTNEDQRRH 162
         || ::|                  |||:|  || |   ::| |: ||
Db   621 TAKENEKKAKSLEADLMQLQEDLAAAERARKQADLEKEELAEELASSLSGRNALQDEKR 680

Qy   163 IEIRDQALSNAQAKVVKLEEELKKKQ-----VYVDKVEKMQQALVQL--QAACEK-REQLE 215
         |||| :  | |  |  :||||||||       | : | :|      |     |  ||:||
Db   681 LE---------ARIAQLEELEEQGNMEAMSDRVRKATQQAEQLSNELATERSTAQKN 730

Qy   216 HRLRTRLERELESLRIQORQGNCQPTNVSEYNAAALMELLREKEERILALEADMTKWEQK 275
         |: || | | :   |   ||    |  | :         |   | :| :| |  |:|:|
Db   731 ESARQOLERONKELR--------SKLHEMEGAVKSKF---KSTIAALEAKIAQ- 773

Qy   276 YLEENVMRHFALDAAATVAAQRDTTVISHSPNTSYDTALEARIQKEEEEILMANKRCLDM 335
         | :|    |: |  |         |:|:| :        |    | | |    ::| |
Db   773 -LEEQVEQEAREKQAATKSLKQK----------------DKKLKEILLQVEDERKMAEQYKEQA 819

Qy   336 E---GRIKTLHAQIIEKDAMIKVLQQRSRKEPSKTEQLSCMRPAKSLMSISNAGSGLLSH 392
         :   ||  ||        | |: | |
Db   820 EKGNARVKQLKRQLEEAEEESQRINANRKLQRELDEATESNEA------MGREVNAL 871

Qy   393 SSTLTGSFIME 403
         |: |:| |  |
Db   872 KSKLRGPPPQE 882
```

RESULT 12
US-08-533-306A-4
Sequence 4, Application US/08533306A
Patent No. 5837457
GENERAL INFORMATION:
;   APPLICANT: Liu, Pu
;   APPLICANT: Collins, Francis S.
;   APPLICANT: Siciliano, Michael J.
;   APPLICANT: Claxton, David

RESULT 13

US-08-533-306A-6
Sequence 6, Application US/08533306A
Patent No. 5837457
GENERAL INFORMATION:
APPLICANT: Liu, Pu
APPLICANT: Collins, Francis S.
APPLICANT: Siciliano, Michael J.
APPLICANT: Claxton, David
TITLE OF INVENTION: Markers for Detection of Chromosome 16
TITLE OF INVENTION: Rearrangements
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:
ADDRESSEE: Harness, Dickey & Pierce, P.L.C.
STREET: P.O Box 828
CITY: Bloomfield Hills
STATE: MI
COUNTRY: USA
ZIP: 48303
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/533,306A
FILING DATE: September 25, 1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Smith, DeAnn F.
REGISTRATION NUMBER: 36683
REFERENCE/DOCKET NUMBER: 2115-00869COB
TELECOMMUNICATION INFORMATION:
TELEPHONE: (810) 641-1600
TELEFAX: (810) 641-0270
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 816 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-533-306A-6

Query Match
Best Local Similarity   6.18;   Score 203; DB 2; Length 816;
                        23.4%; Pred. No. 5.5e-06;
Matches 101; Conservative 74; Mismatches 150; Indels 106; Gaps 16;

```
Qy   30 EILSDENRNLR-------QELEGCYEKVARLQKVETE-----IQRVSEAYENLVKSSSKREA  79
           |: | |||  ||        | |:  || : ||  : :           | :| :  :| ||
Db  432 ELQASEDAKLRLEVNMQALKGQFERDLQARDEQNEEKRRQLQROLHEYETELEDDRNERA 491

Qy   80 LEKAMRNKLEGEIRRMHDFNRDLRERLETPANKQLAE------KEYEGS-EDTRKTISQLFA 133
             | :||||| : | |     |  |||| |: |          |:  | : ||  :|| :||
Db  492 LAAAAKKLEGDLKDLELQADSAIKGREEAIKQLRKIQAQMKDFQORELEDARASRDEIFA 551

Qy  134 KNKESQR------EKEKLEAELATA---RSTNEDQRRH 162
           |:|:          |||:|| || :        |:||||:||
Db  552 TAKENEKKAKSLEADILMQLQEDLAAAERARKQADLEKEELAEELASSLSGRNALQDEKRR 611

Qy  163 IEIRDQALSNAQAKVVKLEEELKKKQ----VYYDKVEKMQQALVQL--QAACEK-REQIE 215
           |:|      : ||:||||:::|            |:|| |||  ||  |  | :| ||
Db  612 LE-------ARIAQLEEELEEOGNMEAMSDRVRKATQOAEQLSNELATERSTAQKN 661

Qy  216 HRLRTRLERELESLRIQORGNCQPTNVSEYNAAAIMELLREKEERILALEADMTKWEQK 275
             ||||:          :| |   :
Db  662 ESARQOLERQNKELR----------SKLHEMEGAVKSKF---KSTIAALEAKIAQ---- 704

Qy  276 YLEENVMRHFALDAAATVAAQRDTTVISHSPNTSYDTALEARIQKEEEITLMANKRCLDM 335
           |||  :       ||:   :       ||: :|:  ||  |
Db  704 -LEEQVEQEAREKQAATKSLKQK--------DKKLKEILQVEDERKMAEQYKEQA 750

Qy  336 E---GRIKTLHAQIIEKDAMIKVLQQRSKEPSKTEQLSCMRPARKSLMSISNAGSGLLSH 392
           |: |: | : |   ||  ::|
```

Db 751 EKGNARVKQLKRQLEEAEEESQRINANRRKLLQRELDEATESNEA-------MGREVNAL 802
Qy 393 SSTLTGSPIME 403
Db 803 KSKLRGPPQE 813

RESULT 14
US-08-742-923A-2
Sequence 2, Application US/08742923A
Patent No. 5869611
GENERAL INFORMATION:
APPLICANT: Liu, Pu
APPLICANT: Collins, Francis S.
APPLICANT: Siciliano, Michael J.
APPLICANT: Claxton, David
TITLE OF INVENTION: Markers for Detection of Chromosome 16
TITLE OF INVENTION: Rearrangements
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:
ADDRESSEE: Harness, Dickey & Pierce, P.L.C.
STREET: P.O. Box 828
CITY: Bloomfield Hills
STATE: MI
COUNTRY: USA
ZIP: 48303
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/742,923A
FILING DATE: No. 5869611ember 1, 1996
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Smith, DeAnn F.
REGISTRATION NUMBER: 36683
REFERENCE/DOCKET NUMBER: 2115-00869DVC
TELECOMMUNICATION INFORMATION:
TELEPHONE: (810) 641-1600
TELEFAX: (810) 641-0270
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 576 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-742-923A-2

Query Match
Best Local Similarity   6.18;   Score 203; DB 2; Length 576;
                        23.4%; Pred. No. 3.5e-06;
Matches 101; Conservative 74; Mismatches 150; Indels 106; Gaps 16;

```
Qy   30 EILSDENRNLR-------QELEGCYEKVARLQKVETE-----IQRVSEAYENLVKSSSKREA  79
           |: | |||  ||        | |:  || : ||  : :           | :| :  :| ||
Db  192 ELQASEDAKLRLEVNMQALKGQFERDLQARDEQNEEKRRQLQROLHEYETELEDDRNERA 251

Qy   80 LEKAMRNKLEGEIRRMHDFNRDLRERLETPANKQLAE------KEYEGS-EDTRKTISQLFA 133
             | :||||| : | |     |  |||| |: |          |:  | : ||  :|| :||
Db  252 LAAAAKKLEGDLKDLELQADSAIKGREEAIKQLRKIQAQMKDFQORELEDARASRDEIFA 311

Qy  134 KNKESQR------EKEKLEAELATA---RSTNEDQRRH 162
           |:|:          |||:|| || :        |:||||:||
Db  312 TAKENEKKAKSLEADILMQLQEDLAAAERARKQADLEKEELAEELASSLSGRNALQDEKRR 371

Qy  163 IEIRDQALSNAQAKVVKLEEELKKKQ----VYYDKVEKMQQALVQL--QAACEK-REQIE 215
           |:|      : ||:||||:::|            |:|| |||  ||  |  | :| ||
Db  372 LE-------ARIAQLEEELEEOGNMEAMSDRVRKATQOAEQLSNELATERSTAQKN 421

Qy  216 HRLRTRLERELESLRIQORGNCQPTNVSEYNAAAIMELLREKEERILALEADMTKWEQK 275
           :|||||:          :| |   :||
```

```
Db    422  ESARQQLERQNKELR--------------SKLHEMEGAVKSKF----KSTIAALFAKIAQ---- 464
             |||:|  |    |             ||:|:|:      |||||||||||
Qy    276  YLEENVMRHFALDAAATVAAQRDTTVISHSPNTSYDTALEARIQKEEEILMANKRCLDM 335
             |:|:|  ||  |:  || | :||       ||     || |  |:  |||
Db    464  -LEEQVEQEAREKQAATKSLKQK--------------DKKLKEILLQVEDERKNAEQYKEQA 510
             ||   ||  |:|:|          ||:|:|||||||||||||
Qy    336  E---GRIKTLHAQIIEKDAMIKVLQQRSRKEPSKTEQLSCMRPAKSLMSISNAGSGLLSH 392
             ||||:| |:||||:|:|||:|:||||||||:|||:|:||||||||||||||||||||
Db    511  EKGNARVKQLRKQLEEAEESQRINANRRKLQORELDEATESNEA-------MGREVNAL 562

Qy    393  SSTLTGSPIME 403
             ||||||||||
Db    563  KSKLRGPPPQE 573
```

```
Query Match          6.1%;   Score 203;  DB 2;  Length 885;
Best Local Similarity  23.4%;  Pred. No. 6e-06;
Matches 101; Conservative  74; Mismatches 150;  Indels 106;  Gaps  16;
```

```
Qy    30   EILSDENRNLR----------QELEGCYEKVARLQKVETE-----IQRVSEAYENLVKSSSKREA 79
             |:| || |          || ||  ||  |:|  |  |:  |:|:| |:|:|
Db    501  ELQASEDAKLRLEVNMQALKGQFERDLQARDEQNEEKRRQLQRQLHEYETELEDERNERA 560
Qy    80   LERAMRNKLEGEIRRMHDENRDLRERLETANKQLAE----KEYEGS-EDTRKTISQLFA 133
             | |||| | |:|:||||:::            | ||     ::| |
Db    561  LAAAAKKLEGDLKDLELQADSAIKGREEAIKQLRKLQAOMKDFQRELEDARASRDEIFA 620
Qy    134  KNKESQR-------------------------EKEKLEAELATA--RSTNEDQRRH 162
             ||:::                          |||::|         |: ||:|
```